THE UNIVERSITY *of* EDINBURGH

Audio-visual interactions in manual and saccadic responses

Elena Makovac

Ph.D. thesis

The University of Edinburgh & University of Trieste

2013

Posvećeno mami, tati, baki i Sandru

Abstract

Chapter 1 introduces the notions of multisensory integration (the binding of information coming from different modalities into a unitary percept) and multisensory response enhancement (the improvement of the response to multisensory stimuli, relative to the response to the most efficient unisensory stimulus), as well as the general goal of the present thesis, which is to investigate different aspects of the multisensory integration of auditory and visual stimuli in manual and saccadic responses. The subsequent chapters report experimental evidence of different factors affecting the multisensory response: spatial discrepancy, stimulus salience, congruency between cross-modal attributes, and the inhibitory influence of concurring distractors.

Chapter 2 reports three experiments on the role of the superior colliculus (SC) in multisensory integration. In order to achieve this, the absence of S-cone input to the SC has been exploited, following the method introduced by Sumner, Adamjee, and Mollon (2002). I found evidence that the spatial rule of multisensory integration (Meredith & Stein, 1983) applies only to SC-effective (luminance-channel) stimuli, and does not apply to SC-ineffective (S-cone) stimuli. The same results were obtained with an alternative method for the creation of S-cone stimuli: *the tritanopic technique* (Cavanagh, MacLeod, & Anstis, 1987; Stiles, 1959; Wald, 1966). In both cases significant multisensory response enhancements were obtained using a *focused attention paradigm*, in which the participants had to focus their attention on the visual modality and to inhibit responses to auditory stimuli.

Chapter 3 reports two experiments showing the influence of *shape congruency* between auditory and visual stimuli on multisensory integration; i.e. the correspondence between structural aspects of visual and auditory stimuli (e.g., spiky shape and “spiky” sounds). Detection of audio-visual events was faster for congruent than incongruent pairs, and this congruency effect occurred also in a focused attention task, where participants were required to respond only to visual targets and could ignore irrelevant auditory stimuli. This particular type of *cross-modal congruency* was been evaluated in relation to the *inverse effectiveness rule* of multisensory integration (Meredith & Stein, 1983).

In Chapter 4, the locus of the cross-modal shape congruency was evaluated applying the *race model analysis* (Miller, 1982). The results showed that the violation of

the model is stronger for some congruent pairings in comparison to incongruent pairings. Evidence of multisensory depression was found for some pairs of incongruent stimuli. These data imply a perceptual locus for the cross-modal shape congruency effect. Moreover, it is evident that multisensoriality does not always induce an enhancement, and in some cases, when the attributes of the stimuli are particularly incompatible, a unisensory response may be more effective that the multisensory one.

Chapter 5 reports experiments centred on saccadic generation mechanisms. Specifically, the multisensoriality of the *saccadic inhibition* (SI; Reingold&Stampe, 2002) phenomenon is investigated. Saccadic inhibition refers to a characteristic inhibitory dip in saccadic frequency beginning 60-70 ms after onset of a distractor. The very short latency of SI suggests that the distractor interferes directly with subcortical target selection processes in the SC. The impact of multisensory stimulation on SI was studied in four experiments. In Experiments 7 and 8, a visual target was presented with a concurrent audio, visual or audio-visual distractor. Multisensory audio-visual distractors induced stronger SI than did unisensory distractors, but there was no evidence of multisensory integration (as assessed by a race model analysis). In Experiments 9 and 10, visual, auditory or audio-visual targets were accompanied by a visual distractor. When there was no distractor, multisensory integration was observed for multisensory targets. However, this multisensory integration effect disappeared in the presence of a visual distractor. As a general conclusion, the results from Chapter 5 results indicate that multisensory integration occurs for target stimuli, but not for distracting stimuli, and that the process of audio-visual integration is itself sensitive to disruption by distractors.

Sintesi

L'ambiente fornisce informazioni che vengono trasportate dai diversi canali sensoriali. Nonostante nelle prime fasi di elaborazione queste informazioni siano segregate tra loro, il risultato finale della percezione è sempre unitario e coerente. Il processo che porta dall'iniziale divisione dell'informazione tra diversi canali sensoriali a un risultato finale unificato viene denominato *integrazione multisensoriale*(Stein & Meredith, 1993).

I sistemi sensoriali sono stati studiati in modo approfondito ma separatamente, in quanto deputati alla raccolta di specifiche forme di energia fisica (quali la luce per la visione, la pressione cutanea per il tatto, le onde acustiche per l'udito e così via). Solo in anni più recenti l'interesse per la loro unione ha prevalso sull'interesse per i singoli componenti.

Dopo i primi studi pionieristici condotti da Meredith e Stein (1983), l'interesse per questo campo di ricerca è cresciuto incessantemente. La ragione è che l'integrazione multisensoriale è la condizione più naturale per l'organismo, e sono rare le volte in cui un oggetto o evento ambientale stimola un solo canale sensoriale, e lascia intatti gli altri.

L'integrazione multisensoriale viene misurata valutando il miglioramento della risposta a uno stimolo multisensoriale in relazione a una risposta agli stimuli unisesensoriali. La misura presa in considerazione è quindi *il miglioramento della risposta multisensoriale* (*multisensory response enhancement*, MRE), che costituisce il principale concetto sul quale si basano gli esperimenti della presente tesi. Il miglioramento viene inteso in diversi modi, a seconda del tipo di studio e dal tipo di compito: nel caso di studi neurofisiologici si tratta di potenziali evocati, negli studi comportamentali invece viene valutata la riduzione dei tempi di reazione o l'aumento di sensibilità.

Tradizionalmente si presumeva che l'integrazione tra le informazioni elaborate indipendentemente dalle diverse modalità avvenisse in aree neocorticali associative 'di alto livello'. Gli studi di Meredith e Stein (1983) invece, hanno individuato il Collicolo Superiore (CS) come una delle principali strutture coinvolte nel processo di integrazione multisensoriale, data la alta percentuale di neuroni multisensoriali presenti nei suoi strati profondi. Gli stessi autori hanno enucleato tre regole dell'integrazione multisensoriale: la *regola spaziale*, la *regola temporale* e *la regola dell'efficacia inversa*. Le regole spaziale e

temporale si riferiscono al miglioramento dell'integrazione per stimoli che accadono in prossimità spazio-temporale. La regola dell'efficacia inversa, invece, riguarda la salienza degli stimoli che devono venire integrati: più gli stimoli sono deboli, più forte è la loro integrazione.

L'obiettivo generale di questa tesi è lo studio di diversi aspetti dell'integrazione tra stimoli uditivi e visivi, nelle risposte sia manuali sia saccadiche.

Dopo l'introduzione generale contenuta nel Capitolo 1, nel Capitolo 2 viene trattato l'effetto della discrepanza spaziale sull'integrazione multisensoriale mediata dal CS. In questo capitolo vengono utilizzati due metodi che permettono di isolare l'attività del CS e di concludere se l'effetto studiato sia mediato da questa struttura oppure no.

Questi metodi si basano su dati neurofisiologici che riportano una sostanziale assenza di afferenze dai coni S (specializzati per le lunghezze d'onda corte) al CS, che sarebbe invece attivato dal segnale di luminanza, cui non contribuisce l'output dei coni S. I dati riportati nel secondo capitolo indicano che la regola spaziale si applica solo agli stimoli che attivano il CS e non a quelli che non lo attivano.

Nel Capitolo 3 vengono presi in considerazione altri fattori che influenzano l'integrazione multisensoriale: la congruenza tra vari attributi degli stimoli, ovvero la *congruenza crossmodale*. Sono stati presentati stimoli "spigolosi" e "rotondi", sia visivi che uditivi, in condizioni congruenti o incongruenti. È stata inoltre manipolata la luminanza degli stimoli visivi, per studiare la relazione tra la congruenza cross-modale e la legge dell'efficacia inversa. Dai dati emerge una superiorità degli stimoli congruenti, soprattutto a livelli di luminanza alta, dimostrando che la congruenza tra attributi e l'efficacia inversa sono effetti tra loro indipendenti.

Nel Capitolo 4, la congruenza cross-modale viene analizzata con l'analisi del race model, che permette di individuare se il meccanismo sottostante l'interazione sia di tipo indipendente (*race model*; Raab, 1962) o interattivo (*co-attivazionale*; Miller, 1982). I dati dimostrano una maggiore violazione del modello per le coppie di stimoli congruenti. In alcuni casi, inoltre, l'accoppiamento di stimoli incongruenti porta a una depressione della risposta. Questo fatto dimostra che non tutti gli accoppiamenti portano sempre ad un MRE, e che a volte la congruenza tra attributi strutturali (come la *forma*) influenza l'integrazione a livello percettivo e non decisionale.

Il Capitolo 5 è centrato sull'integrazione audio-visiva nei movimenti saccadici. In particolar modo, viene analizzata una possibile natura multisensoriale dell'effetto di inibizione saccadica (Reingold & Stampe, 2002). Con il termine *inibizione saccadica* viene indicata la cancellazione delle saccadi verso un target, quando uno stimolo distrattore viene presentato contemporaneamente (o dopo un ritardo temporale) al target. Questo fenomeno, studiato in ambito visivo, è stato investigato nel Capitolo 5 utilizzando anche target e distrattori uditivi e audio-visivi. I risultati indicano che sia il distrattore uditivo che quello audio-visivo producono inibizione, e che l'inibizione con il distrattore audio-visivo è più forte dell'inibizione con distrattori unisensoriali (audio o visivi). E' stato inoltre ipotizzato che un target audio-visivo possa essere più resistente a un distrattore visivo. I dati, però, indicano che l'inibizione è uguale per tutti i tipi di target. L'analisi del *race model* ha evidenziato, a sorpresa, una violazione solo per la condizione in cui non era presente il distrattore. La presenza del distrattore ha, quindi, distrutto i processi di integrazione multisensoriale. L'ultimo capitolo della tesi riporta una conclusione generale e possibili future linee di ricerca.

Acknowledgements

I would like to express my deep gratitude to my supervisor and friend, Professor Walter Gerbino for his patient guidance, enthusiastic encouragement and useful insights for this research work during the last four years.

I would also like to thank my second supervisor, Dr. Rob McIntosh, for his efficient supervision during my staying at the University of Edinburgh, and for being an example to follow.

My deep gratitude goes also to Antimo Buonocore, for introducing me to the world of saccades and for supporting and tolerating all my research related ups and downs.

Thanks to all of my closest friends (to name some of them, Alessandra, Margherita, Nicol, Danijela, Eli) and a special thanks to Chai, for being always there for me.

Un ringraziamento alla mia famiglia, mamma Marija, papà Danilo e Sandro, per essere il vento nelle mie vele ed il mio porto sicuro.

Declaration

I declare that this thesis is my own composition, and that the material contained in it describes my own work. It has not been submitted for any other degree or professional qualification. All quotations have been distinguished by quotation marks and the sources of information acknowledged.

Elena Makovac

Acronyms and abbreviations

AES: anterior ectosylvian sulcus

ANOVA: analysis of variance

AV: Audio-visual

BSBG: *brainstem saccade burst generator*

CMC: Cross-modal congruency

CPD: Cumulative probability distribution

dB: Decibel

DLSC:deep layers of the SC

DP: different position

FEF: frontal eye field

IER: Inverse effectiveness rule

IML: internal medullary lamina

JND: Just noticeable difference

K: koniocellular

LGN: lateral geniculate nucleus

LIP: lateral intraparietal area of the posterior parietal cortex

LLBN: long-lead burst neurons

LO: Lateral occipital

M: magnocellular

MI: Multisensory integration

MLBN: medium-lead burst neurons

MRE: Multisensory response enhancement

MS: Multisensory

MSC: Multisensory congruent

MSI: Multisensory incongruent

OMN: oculomotor neurons

OMR: oculomotor reflex

OPN: omnipause neurons

OR: orienting reflex

P: Parvocellular

RDE: Remote distractor effect

rLS*:* Rostrallateral suprasylvian sulcus

RM: Race model

RMI: Race model inequality

RT: Reaction times

RTE: Redundant target effect

SC: Superior Colliculus

SEF: Supplementary eye fields

SI:Saccadic inhibition

SIH: Saccadic inhibition hypothesis

SNR: Signal-to-noise ratio

SOA: Stimulus onset asynchrony

SP: Same position

SRT: Saccadic reaction times

STS: Superior temporal sulcus

US: Unisensory

USA: Unisensory auditory

USV: Unisensory visual

Table of Contents

Figures

Tables

CHAPTER 1

GENERAL INTRODUCTION

Chapter 1:
General introduction

1.1 The senses

It is not easy to define and classify human senses. Beside the “traditional” sight, hearing, touch, smell and taste (as described by Aristotle) there are many other things that a human being can sense: like pain (nociception); balance (equilibrioception); proprioception; the sense of time; the change of temperature (thermoception) etc.

During decades, researchers studied all of them in a parallel way. Fodor (1983) stated in his *modular theory* that sense modalities are processed independently and that perceptual processes are simple computations for perceptual inferences about the environment. He hypothesized the sense modalities as being modular, separated into parallel processing channels that are not affected by higher level cognitive processes called 'central processes'. Fodor highlighted the functional independence of each of the senses and asserted that perception follows a feed forward information processing scheme only.

A reason to think about the senses as being independent and modular is the fact that there are many differences among them. The first obvious difference is that each of the basic senses uses different sensory receptors to collect physical energy from the world. The information collected is different in each sense: optic information for vision, air pressure for audition etc… The second difference concerns the fact that each sensory modality has its own unique subjective impressions or “*qualia*” (for example, perceived colour is unique to the visual system; pitch is specific to the auditory system etc…). Despite this initial difference, however, the final stage of sensory processing is similar among all the senses: for example, at the central level they are all organized in topographic maps to represent how the information is configured in the environment. The reason of such similarities is that the information is at the end always combined into a unitary percept. The brain

constantly processes, analyse and combine sensory information from different sources, with a final result of a coherent interpretation of the world.

1.2 Multisensory Integration

In a recent few decades the researchers have focused the attention on how are the multiple senses coordinated in a coherent perception, instead of studying them separately. The prevailing view was that the senses were all processed independently, with sensory "binding" occurring in later stages of the processing (Treisman & Gelade, 1980) in later association areas of the brain (Felleman & Van Essen, 1991). In our everyday living, however, it rarely happens that an object or event stimulates only one sensory modality. If we think of a very frequent event, verbal communication, lip movements and speech sounds are two connected inputs that are perceived separately only in particular situations.

From an ecological point of view, having multiple senses provides many advantages: each sense is optimally functional in different circumstances and conveys a different kind of information from the external world, enhancing an organism's likelihood of survival. Moreover, the evolutionary importance of multiple sensory modalities is evident when we know that the brain integrates the information from each sense to optimize the level of perception. A full view of how the senses interact in the brain is important since in a natural environment each of them is hardly stimulated one-by-one separately. Today, on the basis of behavioural, neuropsychological and neuroimaging studies we know that the integration among senses occurs earlier that it was thought.

The term *Multisensory Integration* (MI) denotes the processing of information coming from different modalities (that are potentially related to the same object) into a unitary percept. The MI is part of the more general *binding problem,* of how features or stimuli that are initially represented separately can be paired as belonging to the common object. The MI has been initially studied at neuronal level and later explained at the behavioural level. In the next paragraph, I will describe the neural principles followed by the literature on behavioural data.

Nowadays, the knowledge about MI is influencing many fields: technology development, commercial product ergonomic and clinical assessment and rehabilitation of brain damaged and psychiatric patients. For example, the impact of multisensory targets and distractor during driving has become an important field of application for what regards the safety, comfort and aesthetic (e.g. Hoggan & Brewster, 2007; Ho, Tan & Spence, 2008). Moreover, the improving of computer interfaces is strictly linked to the MI research too (Frisoli & Camurri, 2006).

The perception of commercial products is another field in which the knowledge about MI can be applied. The subjective opinions of users on the quality, robustness, ergonomics, appearance etc. are influenced by the multisensory stimulations (e.g. visual-haptic; Vergara, Mondragon, Sancho-Bru, Company, & Agost, 2011).

In conclusion, MI is crucial for the assessment and rehabilitation of neuropsychological and psychiatric disease. A deficit of MI is evident both in autism (Iarocci & McDonald, 2006; Oberman & Ramachandran, 2008; van der Smagt, van Engeland, & Kemner, 2007) and schizophrenia (de Jong, Hodiamont, & de Gelder, 2010; Ross et al., 2007). In the rehabilitation of neuropsychological patients, multisensory stimulation provides a mechanism for permanent recovery from sensory and spatial deficit, since it modulates the activation of cortical areas in a short and long-term way (Bolognini, Frassinetti, Serino, & Ladavas, 2005; Bolognini, Rasi, Coccia, & Ladavas, 2005; Frassinetti, Bolognini, Bottari, Bonora, & Ladavas, 2005).

The present thesis will focus exclusively on the interaction between vision and audition, although MI has been described also between taste, smell and touch (e.g. Holmes & Spence, 2004; Small et al., 2004; Soto-Faraco, Spence, & Kingstone, 2004; Spence, Pavani, & Driver, 2004; Spence & Walton, 2005).

1.3 The neurophysiology of Multisensory Integration

Stein and Meredith were the pioneers in the study of the interactions among senses in different species, using single-cell recordings with cats (Meredith & Stein, 1986a, 1986b; Wallace, Meredith, & Stein, 1998) and monkeys (Bell, Corneil, Meredith, & Munoz, 2001; Wallace & Stein, 1996; Wallace, Wilkinson, & Stein, 1996). Their research was mostly focused on the role of the Superior Colliculus (SC).

The SC is part of the tectum, located in the midbrain, superior to the brainstem and inferior to the thalamus (Figure 1.1).

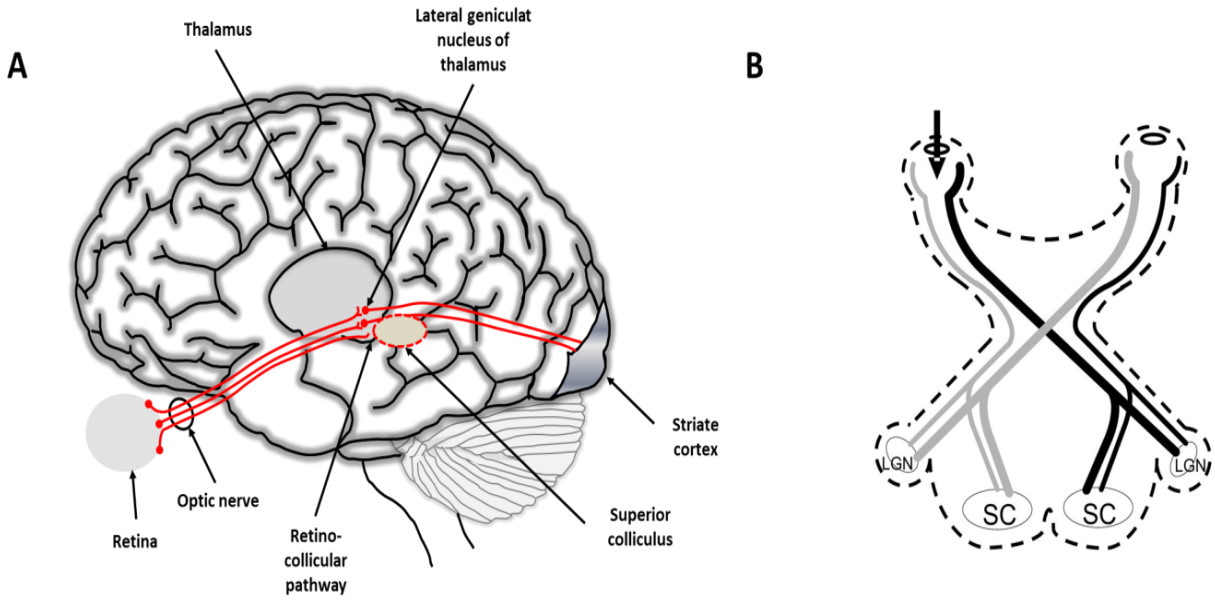


Figure 1.1 Superior Colliculus. A) Representation of the anatomical position of the SC in the brain. B) Two main retino-cortical pathways: retino-thalamic, projecting to the LGN of the thalamus and the retino-tectal, projecting to the SC; both ending into the striate cortex.

The primary function of the SC is to direct behaviour toward objects, even in the absence of the cerebral cortex. For example, cats with major damage to the visual cortex cannot recognize objects, but are still be able to follow and orient toward moving stimuli,

although more slowly than usual (Sprague, 1996). This function involves directly the eye movements. The role of the SC in the generation of saccades will be discussed in Chapter 5.

From top to bottom, the SC is composed of seven cellular layers (Kanaseki & Sprague, 1974), but the structure is usually divided operationally into three parts: the *superficial*, the *intermediate* and the *deep* layers.

The *superficial layers* receive visual information primarily from two sources, i.e., the retina (retinocollicular pathway) and the visual cortex (area 17). Neurons in this layer have receptive fields that respond to contralateral visual stimulation and their activity is primarily related to vision, and the output is mainly directed to the thalamus (Sparks & Hartwich-Young, 1989).

In contrast to the exclusively visual nature of the superficial layers, the *intermediate* and *deep* layers receive projections from many functionally different areas of the brain. The inputs are both “motor” and “sensory”, and its outputs are directed to areas of the brain steam and spinal cord that are involved to the movement of peripheral sensory organs. Their main function is, therefore, the transformation of incoming sensory information into motor commands. In deep layers of the SC receptive fields from somatosensory, visual and auditory modalities converge to form a two dimensional multisensory map of the external world. Here, objects straight ahead are represented caudally and objects on the periphery are represented rostrally. Similarly, locations in superior sensory space are represented medially and inferior locations laterally (Stein & Meredith, 1993).

In my thesis, two main functions of the SC will be investigated: the multisensory integration and the saccadic eye movements. The first one will be discussed in Chapter 2 while the second one in Chapter 5. On a more general note, the rules of the multisensory integration described on the basis of the data observed in the SC, will be applied also for other types of cross-modal processing, presumably non mediated by this structure (e.g. *Inverse effectiveness rule* in Chapter 3 and the *superadditivity* in Chapter 4).

1.3.1 Multisensory integration in SC neurons

The SC has been of interest because of the high proportion of multisensory neurons in its deep layers. The *multisensory neurons* are neurons that respond to stimuli from more than one modality, and they constitute almost the 85% of the neuronal cells of the SC (Fig. 1.2, Wallace et al., 1998).

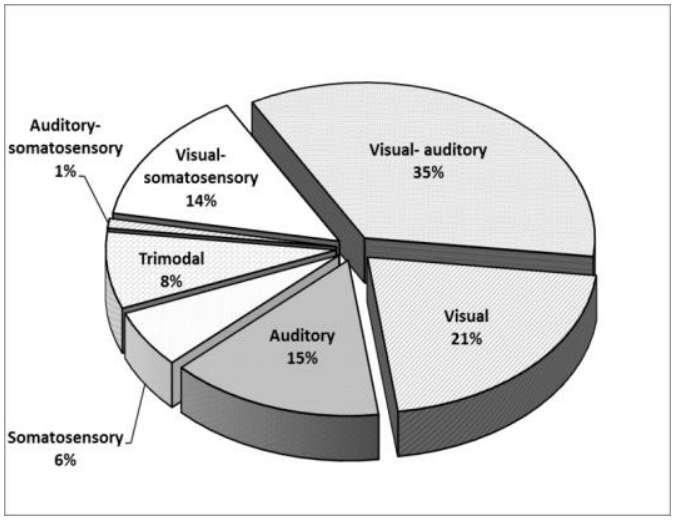


Figure 1.2 *From Wallace et al. (1998).* **Proportion of unisensory and multisensory neurons in the Superior Colliculus.**

The response of the SC multisensory neurons to multisensory stimuli has some peculiar characteristics that distinguish them from the responses to the unisensory components alone (for a review on this early literature see the volume by Meredith and Stein, 1993). At the time of the first studies, it was acceptable to believe that the algorithm employed was the winner-take-it-all, and that MS neurons only characterize a common route for different stimuli to access the same motor program. If that were true, then there were no reasons to believe that the response to a MS stimulus would be qualitatively different from a response to a unisensory stimulus. With the beginning of single-cell recording studies (Meredith & Stein, 1983), a defining criterion for MI was selected: the response to a MS stimulus should be significantly different from the response

to a unisensory stimulus. In neurophysiological studies of the parameter compared was the number of impulses evoked, but the same principle was later applied also on other measures, including behavioural data (RT). After adopting this criterion, it becomes evident that the responses to MS stimuli were often different from the responses to US stimuli, either stronger or weaker. In a recent review, Stein, Stanford, Ramachandran, Perrault, and Rowland (2009) have summarized all the possible computation underlying the MI (Figure 1.3).

For the purpose of the present thesis, I will take in consideration the principle of *superadditivity*: the responses to MS stimuli are often non-linear, i.e. the MS response exceeds the sum of the unisensory responses.

The term *multisensory response enhancement* (MRE) refers to the augmentation of the response of MS neurons, and occurs whenever the combined stimulus response significantly exceeds the best modality-specific response (Meredith & Stein, 1983). More generally, when multiple stimuli are combined an improvement of the MS response (measured as response reliability, number of impulse evoked, duration of the discharge train, etc.) over unisensory is evident, and this is true for every MS combination (visual-auditory, somatosensory-visual, trimodal etc.) (Meredith & Stein, 1986b).

To assess the magnitude of the response enhancement and to evaluate a statistically significant increase in relation to the US response, a simple formula was developed:

$$(CM\text{-}SM^{max})/SM^{max} \times 100$$

where CM is the number of impulses evoked by the MS stimulus and SM_{max} is the number of impulses to the most effective US stimulus (Meredith & Stein, 1983). This enhancement is particularly evident when neither stimulus alone is capable of producing action potentials, yet combined together can evoke a neuronal response. In some neurons, however, the opposite effect can be observed. When combined together, MS stimuli can produce a *multisensory response depression*: i.e. MS stimuli evoke less impulses, shorter discharge train or lower peak frequencies than the two stimuli separately. It has been proved that both response enhancement and depression depend on the spatial and temporal arrangement of the stimuli.

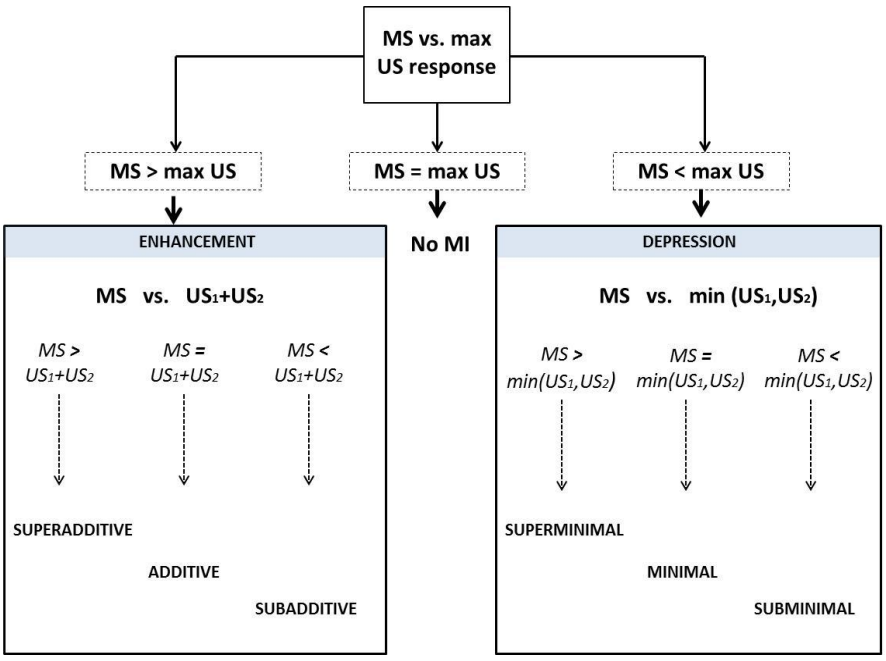


Figure 1.3 *From Stein et al. (2009).* **A schematic representation for evaluating the computation underlying the multisensory integration.** The mean MS response is compared to the most effective US response (max US response). If the MS is larger than the max US, the MS enhancement occurs, if the MS is smaller than the max US, the MS depression occurs. If the MS is equal to the mx US, no MI is identified. The MS enhancement can be further explained by *superadditive, additive and subadditive* processes. The MS depression can be divided into *superminimal, minimal and subminimal*, by comparing the MS response to the less effective US response (i.e. min US).

Whether the multisensory stimuli will result in a response enhancement or depression depends on three factors named the *spatial, temporal and inverse effectiveness rules.*

1.3.1.1 Spatial rule. The spatial rule is particularly crucial for the orienting role of the SC. The SC contains multisensory neurons that are created by the convergence of modality-specific afferents arising from a variety of sources (Wallace, Meredith, & Stein, 1993). As a consequence of this convergence, each multisensory SC neuron has multiple receptive fields, one for each sensory modality to which it responds. These receptive fields represent similar regions of sensory space. For example, a visual-auditory neuron with a visual receptive field in central space will have an auditory receptive field in an overlapping region of central space (Jay & Sparks, 1987; King & Hutchings, 1987; Meredith & Stein,

1996). The cross-modal register of receptive fields is critical for normal multisensory processes (Meredith & Stein, 1986a, 1986b, 1996; Wallace & Stein, 1996). Under normal circumstances, multiple information events will originate from the same location in space and will fall simultaneously within a given neuron's overlapping receptive fields. The spatial rule of MS integration states that spatially coincident stimuli produce MRE (figure 1.3), whereas spatially disparate stimuli produce either depression or no interaction (Meredith & Stein, 1983). When both stimuli fall within the inhibitory region of their respective receptive fields they are not integrated and a neuron's response is no different from the response to the single-modality stimuli presented alone.

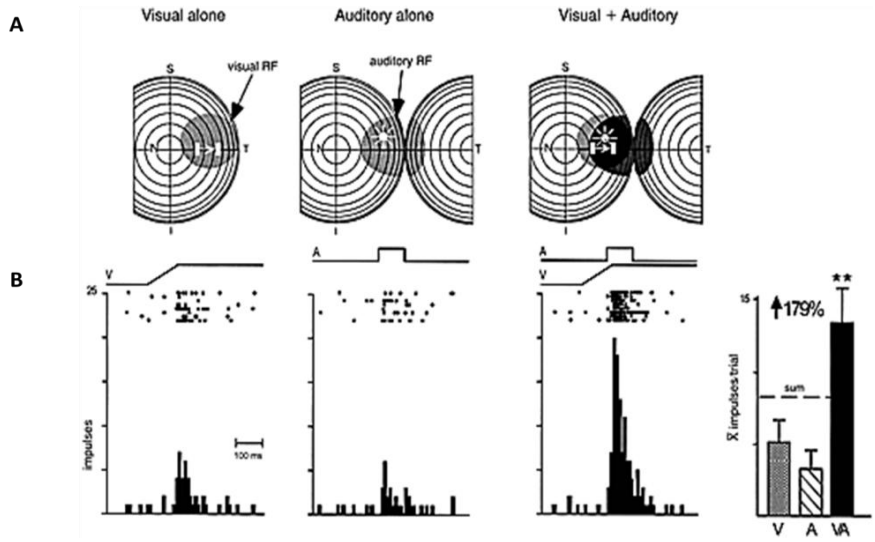


Figure 1.4 *From Meredith and Stein (1983).* A) **Receptive fields of a visual, auditory and overlapping audio-visual stimuli**. B) evoked responses for the visual and auditory stimuli alone and for their AV combination. As evident, the result of overlapping receptive fields is super-additive enhancement in the number of impulses evoked in the AV condition (+179%).

The receptive field is defined as the region in which a sensory stimulus produces excitation. However, the excitatory zone is sometimes surrounded by an inhibitory area. The stimulation of the inhibitory area, in case of two spatially misaligned stimuli, may result in a generalized inhibition and suppress all excitatory inputs to that neuron. Since the receptive fields have to be in register with each other, the SC adopted the solution of linking all the receptive fields from different modalities to the position of the eye. For

example, any time the eye is moved toward a certain direction, it produces a compensatory shift in both auditory and somatosensory receptive fields (Groh & Sparks, 1996), creating a common oculocentric coordinate frame.

1.3.1.2 Temporal rule. In order to be integrated, multisensory stimuli have to be linked also in time (Meredith, Nemitz, & Stein, 1987). The temporal rule of MI may seem intuitive: stimuli occurring at the same time are processed as originating from the same event, while stimuli separated by a time interval are perceived as being two different events. The MI between stimuli occurs over a large "time-window" of integration, but the amount of temporal discrepancy affects the strength of the integration, being maximal when the peak period of activity of the two stimuli coincides. The "time-window" seems a fundamental prerequisite for the integration if we take into account the differences in the energy conveyed to different sensory organs: the mechanical transduction of sound waves in the ear takes less time than the chemical transduction of light at the retina. For example, it takes approximately 13 ms for an auditory stimulus close to the ear to activate the SC, whereas it takes from 65 to 100 ms for a visual stimulus to reach the same neuron (King & Palmer, 1985). The time-window of integration will be discussed in the following paragraph also for behavioural measures and in Chapter 5 for saccade generation.

1.3.1.3 Inverse effectiveness rule (IER). The magnitude of the MRE has been shown to depend not only on temporal and spatial co-occurrence, but also on the characteristics of the unisensory components. Meredith and Stein (1983) reported that the enhancement is stronger with less effective stimuli. In the most extreme situation, two stimuli that do not produce a response in US conditions become effective when combined together. For multisensory depression, the opposite situation is true: the weaker is the unisensory stimulus, the stronger is the depression, represented by a negative MRE (Jiang & Stein, 2003). The two forms of IER must depend on different mechanisms, since they are caused by different stimulus interactions: enhancement through MS co-occurrence and depression through spatial and/or temporal disparity (see also Kadunce, Vaughan, Wallace, Benedek, & Stein, 1997). The majority of studies have investigated the effect by manipulating the salience of the US components, and the results were most of the time coherent with the

general rule: the MRE decreased as stimulus salience increased (Alvarado, Stanford, Vaughan, & Stein, 2007; Alvarado, Vaughan, Stanford, & Stein, 2007; Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Kadunce et al., 1997; Meredith & Stein, 1986a, 1986b; Stanford, Quessy, & Stein, 2005). In other words, as the information from a US stimulus increases (e.g., by increasing its saliency and therefore amplifying its efficiency), the amount of "error" resolved by adding a non-informative source decreases, resulting in a proportional benefit. The study by Stanford et al. (2005) was the first systematic and quantitative assessment of the integrative modes of SC multisensory neurons. The authors recorded the single-neuron responses in the cat SC, for auditory, visual and audio-visual stimuli, over a large range of stimulus intensities. The visual stimuli ranged from 0.65 to 13.0 cd/m^2 and auditory stimuli from 0.7 to 70 dB. The authors concluded that the efficacy of the US inputs is the main factor dictating whether the multisensory computation is superadditive, additive, or subadditive. In particular, superadditivity was most prevalent with weakly efficient US stimuli. The results were additive with less efficient stimuli, and in some cases even subadditive, when most effective stimuli were combined.

While the result at the neuronal level seems to be clear, the behavioural evidence of the IER is less clear, and sometimes even contradicts the rule. The behavioural data will be discussed further in Chapter 3 and in the General conclusion chapter.

1.3.2 Multisensory Integration in cerebral cortex

Even if the SC is the most investigated structure for the MI, recent neurophysiological and neuroimaging studies have found that MI is under the influence of the cortex as well.

Multisensory neurons of the SC receive inputs from the association cortex (Wallace & Stein, 1994), and its orientation behaviour are mediated by two cortical areas—the *anterior ectosylvian sulcus* (AES) and the rostral aspect of the *rostral lateral suprasylvian sulcus* (rLS; Jiang, Jiang, & Stein, 2002; Jiang et al., 2001). The AES contains somatosensory (Burton & Kopf, 1984; Clemo & Stein, 1982, 1983), auditory (Clarey & Irvine, 1986), and visual (Benedek, Mucke, Norita, Albowitz, & Creutzfeldt, 1988; Olson & Graybiel, 1987) projecting to multisensory neurons of the SC (Meredith & Clemo, 1989; Wallace et al., 1993) and determining the product of the MI.

In cats, for example, multisensory orientation behaviours are compromised when the AES is deactivated (Wilkinson, Meredith, & Stein, 1996). Moreover, the animal can still use modality-specific cues to detect and locate a target, but they do not exhibit any multisensory enhancement in the presence of multisensory cues.

In humans, Alvarado et al. (2007) outlined a contrast in the influence of these areas over the integration of multisensory (visual–auditory) stimuli and the absence of such influence over the integration of within-modal (visual–visual) stimuli. Their deactivation is in direct proportion to the magnitude of multisensory enhancement, whereas no effect on responses to stimulus combinations that do not yield multisensory enhancement. These data indicates a specific corticocollicular mechanism that modulates the magnitude of response enhancement mediated by AES/rLS.

Stein & Wallace (1996) described many similarities between the MS neurons in the SC and those in the AES of the cat: they have multiple receptive fields for different sensory inputs, and these receptive fields overlap one another.

Other association areas are considered as being multisensory: human superior temporal sulcus (for a review see Beauchamp, 2005; Ghazanfar & Schroeder, 2006) and lateral intraparietal area of the posterior parietal cortex (LIP). Although LIP was for a long time considered a unimodal visual area, neurons in LIP are now known to be multisensory,

receiving a convergence of eye position, visual and auditory signals (Andersen, Snyder, Bradley, & Xing, 1997). There is growing evidence supporting MI in unisensory areas as well. Neural activity in the visual cortex (measured with fMRI) is stronger when a light source is flashed near the tip of the index compared to the light stimulus alone (Macaluso, Frith, & Driver, 2000). Morrell (1972) already showed that 41 % of visual neurons can be driven by auditory stimuli showing a degree of spatial tuning. Moreover, there are evidence of multisensory activity in other unisensory areas such as auditory and somatosensory cortex (Ghazanfar & Schroeder, 2006).

In conclusion, other areas that have been proved to have multisensory properties (either by anatomical or eletrophisiological data) are also the frontal and prefrontal cortices (e.g. Sugihara et al.,2006), especially in the neurons of the ventrolateral prefrontal cortex of the macaques.

The multisensory rules described for the SC are less clear in cortical areas. For example, the superadditivity has been described sometimes, but it is more the exception that the rule (for a review see Alais, Newell, & Mamassian, 2010). In a fMRI study, Foxe et al. (2002) reported superadditivity for combinations of audio-tactile stimuli in the Superior temporal sulcus (STS). Stevenson and James (2009) varied the saliency of the stimuli and found evidence for superadditive responses to weak stimuli in STS. The general rule, however, is that the MS responses in cortical areas are weaker that those registred at the level of the SC. However, the major difference between the MI in the cortex and in the SC regards the fact that cortical areas are more likely to be influenced by the semantic (or featural) congruency of the stimuli. The congruency between the stimuli (and the areas involved) will be further discussed in Chapters 3 and 4.

1.4 Perceptual and behavioural consequences of multisensory Integration

Researchers use the SC as a general model when trying to understand the MI at neuronal, behavioural and perceptual level (Stein et al., 1989). Later studies revealed that the principles of the MI in SC neurons (spatial, temporal and inverse effectiveness rule) and behavioural data in cats (Burnett, Stein, Chaponis, & Wallace, 2004; Burnett, Stein, Perrault, & Wallace, 2007; Jiang et al., 2002; Stein, 1988; Stein, Huneycutt, & Meredith, 1988; Wilkinson et al., 1996) and humans (e.g. Diederich, 2008; Frassinetti, 2005; Bolognini, Rasi, & Ladavas, 2005) are strongly correlated.

The superior colliculus (SC) is believed to play an important role in sensorimotor integration and orienting behavior. Traditionally, it is divided into superficial layers (containing mostly visual neurons, see Chapter 1) and deep layers (containing multisensory and premotor neurons; Tardiff et al., 2005). Superficial layers are connected mainly with structures involved in visual perception, whereas the neurons of the deep layers mostly motor-related (for review see, e.g. Huerta and Harting 1984). The influence of the mutlisensory stimulation on the behaviour initiates in multisensory neurons (by enhancing their firing rate in response to the multisensory stimulus) which influences the level of activation of the premotor neurons in the SC, closely linked to the initiation of a movement. The premotor neurons are facilitated when evoked by multisensory audio-visual stimuli (Peck, 1987), and these cells are the main connection between the sensory enhancement and the motor behaviour mediated by the SC.

The behavioural consequences of the MI are the speeding up of the response (Gielen, Schmidt, & Van den Heuvel, 1983; Hershenson, 1962; Todd, 1912) and the improving of the stimulus detection (decreasing the sensory uncertainty), both in manual responses and in saccadic movements (Diederich & Colonius, 2008). For the purposes of this chapter, only the data on manual responses will be reported, while the MI in saccadic movements will be discussed in Chapter 5. On a more general note, the MI has been found to change the qualitative experience of the sensory percept (McGurk & MacDonald, 1976;

Shams, Kamitani, & Shimojo, 2000), giving rise to the so called multisensory illusions (discussed in next paragraphs).

1.4.1 The influence of MI on reaction times

The speeding of the response to multiple targets in relation to a single target has been described early in the literature and named the *redundant target effect (RTE)*. This effect has been described by Todd (1912) as:

> *"[...] the more rapid responses to the simultaneous stimuli than to the single stimuli... is ... due to a virtual increase in the extensity or intensity of the stimulus...The three simultaneous stimuli summate in excitatory effect and send their discharge down one common tract to the reaction finger" (p. 63)".*

This RTE is evident both in manual and saccadic RTs, comparing multisensory vs. unisensory events (Amlot, Walker, Driver, & Spence, 2003; Diederich & Colonius, 1987; Miller, 1982, 1986) but also with stimuli within the same modality (e.g. Schwarz & Ischebeck, 1994) and for specific populations e.g. hemianopics (Leo, Bolognini, Passamonti, Stein, & Ladavas, 2008); or individuals who have undergone split-brain surgery (Reuter-Lorenz, Nozawa, Gazzaniga, & Hughes, 1995; Savazzi & Marzi, 2004).

In order to avoid semantic confusion regarding the terminology, in the present thesis I will refer to the speeding up of the response in multisensory condition as to the *multisensory response enhancement (MRE)*, and avoid the term redundant target effect, except in the case where it is explicitly required by the literature that I will describe.

Similarly to the MRE at neuronal level, the magnitude of the MRE at behavioural level in the speeding up of the RT is influenced by the temporal and spatial discrepancies between stimuli.

Hershenson (1962) presented visual, auditory and audio-visual stimuli to participants, and asked them to respond as quickly as possible after the detection of any of them. The A and V stimuli appeared at different SOA. The author pointed out that the

integration is not optimal when the A and V stimuli are simultaneous (as suggested by Todd, 1912), but a small SOA equivalent to the difference between the RT to A and V alone is optimal for the MRE to occur.

Two different theories have been proposed in order to explain the speeding of the response in multisensory conditions. In particular, the manner in which the extracted information is combined or integrated to activate the appropriate response has been a matter of some debate. As a result, two classes of models have been advanced as explanations for the effect (for a review see Miller, 1982). One approach is based on the idea that signals on different channels produce separate activations, each of which builds to the level at which it can produce a response. That is, on any particular trial, responding is controlled by the detection of a signal on one channel or the other. The system never combines activation from different channels in order to meet its criterion for responding. Responses to redundant signals are especially fast, according to separate-activation models, because they are produced by the faster of two processes with randomly varying durations. These models are commonly called "*race models,*" because the response to a redundant signal is produced by the winner of the race between two separate response activation processes. On average, the time of the winner will be less than the time for either racer. In view of the latter two assumptions, Raab (1962) summarized race models by saying that they produce statistical facilitation. Alternatives to separate-activation models allow both components of a redundant signal to influence responding on a single trial. It is convenient to characterize these models as allowing activation from different channels to combine in satisfying a single criterion for response initiation, so they are referred to in this article as *co-activation models*. Responses to redundant signals are especially fast, according to the co-activation view, because two sources provide activation to satisfy a single criterion. Activation naturally builds faster, on average, when it is provided by two sources rather than one. This second approach is in accordance to the rules of the MI described at the level of SC neurons. Miller (1982) proposed a direct test of the assumption of separate activation in speeded detection tasks. Since the responses to redundant signals were too fast to be explained as the faster of two responses to individual signals, he assumed that signals jointly contribute to the process of producing the response.

To qualify the nature of the interaction, Miller (1982) developed a mathematical method that allowed the researchers to conclude if the MRE could be explained by probability summation (separate activation or Race model), or neural co-activation. The method compares the probability distribution of reaction times in the multisensory condition to the joint probability distribution of visual and auditory responses, and its assumption is the following: if the probability of responding to the multisensory stimulus is significantly greater at a given time point that the response predicted by the summed probability of responding to each of the unisensory stimuli, a co-activation occurred.

The test makes use of the cumulative probability distributions (CPD) of the latencies obtained for the MS stimuli and for each of the US stimuli. Each participant's CPD are used to calculate the race distribution, according to the following formula:

$$P(AV) \geq [P(A)+P(V)\ ]\ -\ [P(A)\times P(V)]$$

Where P(A) is the probability of responding by a given time on a unisensory auditory trial (A), and P(V) is the probability of responding by a given time on a unisensory visual trial (V), and the last term is the covariance between the separate activations of the two unisensory A and V targets. Miller pointed out that since the last term in the equation is equal to or greater than zero, it follows that with separate activation, for all values of t:

$$P(AV) \geq P(A)+P\ (V)$$

In other words, separate activation models require that the probability of RTs smaller than some value of t obtained with MS stimuli cannot exceed the sum of the probabilities obtained with US stimuli. Violation of this inequality allows us to reject the separate activation models. On the other hand, co-activation, with the assumed pooling of the activation produced by each target occurrence, is consistent with violations, since the RTs to redundant targets can be faster than the fastest response to a single target.

Distributional tests using the race model inequality (see APPENDIX) often show that the observed redundancy gain is actually larger than predicted by the Race model.

In conclusion, these results are in line with the superadditivity of MI assumed also by Stein and Meredith (1993):

> *"[…] The integration of inputs from different sensory modalities not only transforms some of their individual characteristics, but does so in ways that can enhance the quality of life. Integrated sensory inputs produce far richer experiences than would be predicted from their simple coexistence or the linear sum of their individual products"*

1.4.2 The influence of MI on stimulus detection

The MI can enhance perceptual clarity and reduce ambiguity about the stimuli in the environment, improving the detection (Ernst & Bülthoff, 2004; Stein, Meredith, & Wallace, 1993). The speeded detection (and discrimination) of a visual target is faster and often more accurate when an auditory non-target stimulus is presented on the same side as the visual target rather than on the opposite side (e.g. Bolognini et al., 2005; Bolognini et al., 2005; Spence & Driver, 1997).

The first studies on the detection were conducted on animals. Stein, Meredith, Hunneycutt, and McDade (1989) examined the orientation behaviours of cats in response to unisensory and multisensory stimuli. In their study a low intensity visual cue were presented to the animals, and theirs performance significantly improved when a low intensity auditory stimulus appeared simultaneously. Such multisensory enhancement was only evident with auditory stimuli presented at the same location as the visual stimulus. On the contrary, when the auditory stimulus occurred at a different spatial location their performance was significantly depressed (Stein et al., 1989).

Similar results were obtained in later studies with humans as well: an auditory uninformative peripheral cue enhances the perceptual processing of a subsequent visual stimulus (Driver & Spence, 1998; McDonald, Teder-Salejarvi, & Hillyard, 2000; Spence & Driver, 1997). These findings have been interpreted as evidence of the existence of

crossmodal spatial attention: the exogenous shift of attention in one modality (audition) leads to a corresponding shift of attention in another modality (vision; Driver & Spence 1998; Macaluso et al. 2000).

In a speeded detection task, the improvement of the detection can be attributable to post-perceptual factors, since we are more willing to respond (Sperling & Dosher, 1986) or because the uncertainty of the decision is reduced (Shiu & Pashler, 1994), and cannot be taken as an evidence of an improvement at a perceptual level.

In order to overcome this ambiguity, McDonald et al. (2000) used signal detection measures to investigate whether an accessory sound can influence the perception of a visual stimulus at a perceptual or post-perceptual level. The signal detection analysis allowed the authors to separate the two components involved in perception processes: the d′ parameter, which reflects subject's perceptual sensitivity to differentiate a sensory event from its background (perceptual level) and the β parameter, which reflects subject's decision criterion of response (post-perceptual level). The spatial discrepancy between target and accessory auditory cue was manipulated, with the target appearing either on the same position of the cue or controlaterally. Their results showed that the auditory cue influenced the performance both at the perceptual (d') and decisional (β) level, and the enhancement was stronger for the "valid" (same target-cue position) vs. "invalid" (different target-cue positions) trials.

A similar analysis was performed also by Frassinetti, Bolognini, and Ladavas (2002). If multisensory integration facilitates visual perceptual processing, the d' parameter should vary with the degree of spatial and temporal correspondence between the auditory and the visual stimuli, following the spatial and temporal rules (Meredith & Stein, 1986).

On the contrary, if multisensory integration affects post-perceptual decision processes, a reduction of the decision criterion parameter β is expected when a simultaneous sound is presented in the same location. In their study, participants detected below-threshold visual stimuli in US visual or MS audio-visual conditions. The capacity of an acoustic stimulus to enhance the detectability of a visual stimulus was evident only when the two stimuli occurred simultaneously. By contrast when the acoustic stimulus preceded the visual stimulus of 500 ms the visual detectability did not improve. Their result

proved that the sound influenced vision at early perceptual levels more than at later, decision-related levels. Moreover, these results are in line with the temporal rule of MI.
Bolognini et al. (2005) investigated further the influence of both the spatial and temporal disparity on the MRE in a detection task by varying the temporal interval (0, 100, 200, 300, 400, and 500 ms) spatial disparity (0°, 16°, and 32°) between the visual and auditory stimuli. Subjects were required to detect visual stimuli presented below threshold either in unisensory or audio-visual condition, whereas the auditory modality was task-irrelevant. The results confirmed the conclusions by Frassinetti et al. (2002): the detectability (d') increased only when the visual and auditory stimuli were in the same spatial position and simultaneous, and no enhancement was evident when the two stimuli were spatially disparate or at different time delays (auditory stimulus preceded the visual stimulus of 100 – 500 ms).

On the contrary, the response criterion (β) was influenced by the presence of the auditory stimulus, but not in a spatial specific way: β values were lower in all multisensory conditions than in unisensory ones, but independently of both the spatial position of the auditory stimulus and the temporal interval between the two stimuli.

In conclusion, all the described studies provide a solid evidence of a perceptual locus of the audio-visual integration, especially for weakly efficient US stimuli.

1.4.3 Multisensory Illusions

Beside the enhancement and depression of the response, sometimes two different stimuli can interfere with each other's perception, especially in cases of large interstimulus temporal or spatial discrepancies. This particular situation usually results in some distortion of the percept, defined as *multisensory illusions*. These effects are generally explained in relation to two hypotheses: the *modality appropriateness hypothesis* (Welch & Warren, 1980) and the *unity assumption* (Welch, 1999). Since some stimulus proprieties are processed more accurately in one sensory modality than in another, the modality appropriateness hypothesis states that the sensory information with the higher acuity with respect to the relevant task plays a dominant role in the outcome of multisensory

integration (Bonath et al., 2007; Navarra, Alsius, Velasco, Soto-Faraco, & Spence, 2010). For example, the visual system usually dominates audio-visual spatial processes because it has a higher spatial acuity than the auditory system, while the auditory tends to be more influential in terms of temporal analysis. On the other hand, the unity assumption relates to the degree to which an observer infers (not necessarily consciously) that two sensory inputs refer to a single unitary distal object or event.

The ventriloquism-effect (Howard & Templeton, 1966), for example, describes the illusory misallocation of an auditory stimulus towards a simultaneous spatially disparate visual stimulus, and it was suggested to result from 'visual capture' of the speech information of the dominant visual sense. Although the common finding is that the light has an influence on the location of the sound (e.g., Vroomen, 1999), under certain conditions, one can also observe that the sound attracts the location of the light (Bertelson & Radeau, 1981). The spatial attraction thus occurs both ways, and is rather independent of where endogenous or exogenous spatial attention is located (Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001a).

The influence of the sound over the light perception is evident in the double flash illusion (Shams et al., 2000): when a single visual flash is accompanied by multiple auditory beeps, the single flash is incorrectly perceived as multiple flashes.

The McGurk effect (McGurk & MacDonald, 1976) is an illusion that occurs when speech sounds do not match the sight of simultaneously seen lip movements of the speaker, leading to the perception of a phoneme that is different from both the auditory and visual inputs. When listeners hear 'baba' and at the same time see a speaker articulating 'gaga', they tend to combine the information from the two sources into 'dada'. These illusions underscore the strong tendency to bind auditory and visual information that under normal (congruent) circumstances helps reduce stimulus ambiguity.

1.5 Attention and multisensory integration

The role of attention in MI is controversial (McDonald, Teder-Salejarvi, Di Russo, & Hillyard, 2003; Shore & Simic, 2005; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). In general, MI is thought as an automatic process. Other authors, however, argue that the MI is not automatic and can be more or less effective when tasks are performed in multiple modalities (Alsius, Navarra, Campbell, and Soto-Faraco, 2005). Nowadays, we know that attention can influence target selection both through bottom-up (depending on target salience) and top-down (depending on target relevance for the task and voluntary intention of the observer) mechanisms. Given this distinction, the literature provides evidence of a bidirectional influence between MI and attention. For example, spatially and temporally coincident stimuli are more likely to get detected and further processed that the unisensory components alone. This shows that attention tends to orient more easily towards sensory input that possesses multisensory properties, proving that multisensory bottom-up processes can lead to a capture of attention (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; Driver, 1996; Spence & Driver, 2004; Vroomen et al., 2001b). On the other hand, spatial attention has been shown to affect the earliest multisensory components of the event-related potential (Talsma, Doty, & Woldorff, 2007), which at least suggests that attention is involved in MI processes.

In the present thesis the role of the attention will not be directly manipulated. However, most of the results are obtained in a focused attention paradigm, in which the participants should focus their attention towards only the visual modality. Given the distinction between focused attention paradigm and redundant target paradigm (mostly adopted by other studies on multisensory integration), some speculatios will be provided in the general conclusion chapter, mainly as indications for furute investigations aiming to clarify the role of the attention in the effects described in the present thesis.

1.6 Objectives and organisation of the thesis

The general goal of my thesis is to investigate different aspects of the audio-visual interaction in the manual and saccadic reaction times. Specifically, different circumstances have been investigated, under which the MI is favourite or interfered, resulting in a multisensory response enhancement or response depression. In Chapter 1, the influence of spatial coincidence has been investigated with luminance and equiluminant stimuli, using two different methods: the luminance noise and the tritanopic technique (Cavanagh et al., 1987; Stiles, 1959; Wald, 1966). The main conclusion in Chapter 1 is that the spatial rule is valid only under some circumstances (activation of the SC) and invalid when such criteria are not fulfilled. Moreover, we validate the tritanopic technique as a simple and efficient method for the creation of equiluminant stimuli, applied in the study of SC mediated effects. In Chapter 2 the influence of the structural information from the audio and visual stimuli on the MI has been investigated. A congruency effect between auditory and visual modality features has been shown. Moreover, this effect has been proved to arise at a low level of the stimulus elaboration, and the amount of such an effect is strictly related to the salience of the unisensory visual stimuli. In Chapter 3 the shape congruency effect observed in Chapter 2 was tested using the Race model analysis to further investigate whether it arises at a perceptual or decisional level. Chapter 5 is entirely dedicated to the audio-visual interaction in the generation of saccades. The Saccadic inhibition (SI) paradigm, usually adopted in the study of the effect of visual distractors over visually-guided saccades, has been used for the study of the influence of auditory and audio- visual distractors and targets in the process of target selection.

1.7 A brief note regarding the terminology

A brief note regarding the terminology adopted in the present thesis will be provided. The main reason for this clarifying paragraph is the semantic confusion that I have found in the literature regarding different terms sometimes used for the same phenomenon, or the same term used for different phenomena. One of the reasons of this semantic confusion especially in the field of the multisensory integration is perhaps the fast growing of the research in this particular field.

A first basic distinction that will be introduced is the one between the *multisensory* and *unisensory* stimuli. The first ones (also addressed as cross-modal and multimodal stimuli in literature) will denote all those pairings of stimuli coming from different modalities. In some cases, however, the more specific term *audio-visual* will be used, as a subcategory of multisensory stimuli, as opposite to the *unisensory* (US, or *unisensory visual* – USV- and *unisensory auditory* -USA) stimuli. The synonyms *multimodal* and *unimodal*, frequently used in the literature on the multisensory integration, will be here avoided, to prevent semantic confusion.

The second distinction defined is the one between *multisensory processing* and *multisensory integration*. I will adopt the term multisensory processing to describe all those processes that involve more than one sensory modality but not necessarily specifying the exact nature of the interaction between them (Calvert, Spence, & Stein, 2004).

Two subcategories of the multisensory processing that will be discussed in the thesis are the *multisensory integration* and the *cross-modal processing*.

In the literature, both neurophysiologists and behavioural scientists use the term multisensory integration. In the first case, it has a very specific meaning, defined operationally as the augmentation in the number of neuronal impulses to the multisensory stimulation, in comparison to the unisensory stimulations.

In the present thesis, the term will be used to address the neural process by which unisensory signals are combined to form a new product. It is operationally defined as a multisensory response (neural or behavioural) that is significantly different from the responses evoked by the modality-specific component stimuli (see Stein et al., 2009;

Calvert et al., 2004). The difference between the multisensory and unisensory response will be addressed with the term *multisensory response enhancement* (MRE). The amount of MRE (calculated with the formula reported above), will be addressed as the *percentage of multisensory response enhancement* (%MRE), and indicates the proportionate difference between a multisensory response to a multisensory stimulus and the unisensory response to the most effective modality-specific component stimulus. The %MRE is the equivalent of what is called by other authors the *multisensory index* (Stein & Meredith, 1993; Stein et al., 2009). On a more general level, the term multisensory response enhancement will be used to address the general improvement (at the neuronal or behavioural level, both on RTs and detection) in multisensory conditions. The term *redundant target effect*, used in literature to denote the speeding up of the response when multiple stimuli are present (either within- or cross-modally) will be avoided for the explanation of my results.

The second subcategory of the multisensory processing is the *cross-modal processing*. In the present thesis, the term cross-modal processing will refer to all those cases in which different modalities are compared in order to estimate the *cross-modal equivalence or similarities* among them. The effect obtained from the cross-modal processing is the *cross-modal congruency (CMC)*: the superiority of similar (congruent) pairings vs. different (incongruent) pairings in the processing. Such terms have been used to describe the non-arbitrary associations that seem to exist between different basic physical stimulus attributes, or features, in different sensory modalities (e.g. pitch in the auditory and bright in the visual modality), and that influence the processing of the multisensory information.

This is perhaps the term that brings more confusion in the literature. In fact, other synonyms that are frequently used, and that will be avoided in my thesis, are: synaesthetic correspondences (Martino & Marks, 2000; Melara & O'Brien, 1987; Walker et al., 2010), synaesthetic associations (Parise & Spence, 2008), cross-modal equivalences (Lawson & Turkewitz, 1980), cross-modal similarities (Marks, Hammeal, & Bornstein, 1987), and natural cross-modal mappings (Evans & Treisman, 2010; for a recent review see Spence, 2011).

CHAPTER 2

AUDIO-VISUAL INTEGRATION, THE SUPERIOR COLLICULUS, AND THE SPATIAL RULE

Chapter 2

Audio-visual integration, the Superior Colliculus, and the spatial rule

Chapter overview. Chapter 1 provided the current state of research on multisensory integration. Chapter 2 is focused on the *spatial rule* of MI and the involvements of the SC in mediating the effects of spatial correspondence on MI. The spatial rule has been investigated in animals through single cell recording in the SC (Stein & Meredith, 1993). In humans, non-invasive methods have been exploited in order to obtain evidence relevant for inferences about the role of the SC.

The experiments described in Chapter 2 utilise the *S-cone method* (Sumner, Adamjee, & Mollon, 2002), which takes advantage of the lack of projections from short wavelength cones (S-cones) to the SC (Marrocco & Li, 1977) and to the magnocellular pathway in general (Gouras, 1968). In principle, by activating only the S-cones one can avoid the activation of the SC. I compared SC-efficient (red, mediated mainly by L-cones) vs. SC-inefficient (blue, mediated mainly by S-cones) targets, minimizing the luminance contrast of SC-inefficient targets relative to the background, by means of the random luminance modulation technique (Barbur, Harlow, & Plant, 1994; Birch, Barbur, & Harlow, 1992; Mollon, 1982) in Experiments 1 and 2, and the tritanopic technique (Cavanagh, MacLeod, & Anstis, 1987) in Experiment 3. A stronger MRE was evident with spatially coincident vs. spatially non-coincident stimuli only in the case of SC-efficient targets. No spatial rule was evident with SC-inefficient stimuli. Moreover, the MRE in the spatially coincident position was stronger for SC-efficient that for SC-inefficient targets.

I hypothesized that the stronger MRE obtained with SC-efficient targets, being mediated by the SC, should not require the distribution of attention over both sensory modalities. To support such a hypothesis, the three experiments were run in a *focused attention paradigm* in which participants were required to focus on visual targets while ignoring the auditory modality.

2.1 Introduction

2.1.1 A brief introduction to visual pathways

The retina contains two types of photoreceptors: rod and cones. The rods are usually active during low light conditions, whereas the cones are active during daylight vision and carry information about colours and shapes. There are normally three kinds of cones in the human eye, named from the portion of the visible wavelength spectrum they are most sensitive to: short (S), medium (M), and long (L). Sensitivities peak, respectively, at 420–440 nm for S-cones, 534–545 nm for M-cones, and 564–580 nm for L-cones. The photoreceptors project to the retinal ganglions, a type of neurons located near the inner surface of the retina of the eye. The lateral geniculate nucleus (LGN) is the primary relay centre for visual information received from the retina of the eye, and it is found inside the thalamus of the brain. In humans and macaques the LGN has six distinctive layers: two layers, 1 and 2, are called the *magnocellular* layers (MC), while the outer four layers, 3, 4, 5, and 6, are called *parvocellular* layers (PC). Moreover, a set of neurons called the *koniocellular* sublayers (KC), are found ventral to each of the MC and PC layers (Carlson, 2007). The MC, PC, and KC layers of the LGN correspond to the similarly named visual pathways. The MC pathway is believed to receive input from L- and M-cones, but not from S-cones. The MC pathway roughly corresponds to the *luminance pathway* (so called since it mostly carries luminance information, and takes its input from the summed activation of L- and M-cones, although S-cones may contribute to some extent (Drum, 1983; Lee & Stromeyer, 1989; Stockman, Sharpe, Zrenner, & Nordby, 1991). On the contrary, PC neurons are colour-opponent (i.e., their responses are excitatory or inhibitory depending on the stimulus wavelength; De Valois, 1960), thus mediating primarily the perception of colour, by differentiating the outputs of M- and L-cones (King-Smith & Carden, 1976; Sperling & Harwerth, 1971).

The MC pathway has little colour sensitivity and responds preferentially to low spatial and high temporal frequency stimuli, whereas the PC pathway responds to high spatial and low temporal frequency stimuli (Shapley, 1990). The response of the MC

pathway seems to be independent of the wavelength of the light, and this property is referred as broad-band spectral sensitivity (Schiller & Malpeli, 1977). The MC neurons are thought to be 8 times more sensitive to the luminance contrast in comparison to PC neurons (Kaplan & Shapley, 1986).

The greater sensitivity of PC neurons to chromatic contrast (in macaque monkeys and marmosets) and to higher spatial frequencies (in all primate species examined) has been linked to the processing of detail and colour while the greater sensitivity of MC neurons to higher temporal frequencies has been linked to motion perception (DeYoe & Van Essen, 1988; Livingstone & Hubel, 1987; Kaplan & Lee, 1990).

However, the superposition between the MC and the luminance pathways, as well as between the PC and the colour pathways, is controversial (Livingstone & Hubel, 1987; Zeki & Shipp, 1988; Cavanagh, 1991; Li et al., 2007) and a full analysis of this issue is beyond the scope of this thesis.

Stimuli that are uniquely detected by S-cones are processed by a component of the KC pathway projecting from the lateral geniculate nucleus (LGN) to the cytochrome-oxidase blobs of primary visual cortex (V1; Hendry & Reid, 2000). The KC pathway makes up a third functional channel in primate LGN and forms three pairs of layers in macaques, one of which (the middle pair), relays input from S-cones to V1 blobs.

It has been hypothesized that MC and PC neurons support distinct extrastriate visual pathways, based upon differences in physiological signatures and upon their separate projection pathways to V1 and within V1 (Livingstone & Hubel, 1987; Merigan & Maunsell, 1993). One major retino-cortical visual pathway is the retino-collicular pathway, routing primarily through the SC and thalamic pulvinar nucleus onto posterior parietal cortex and visual area MT. Since the SC has been extensively investigated for its multisensory properties, the retino-collicular pathway will be the focus of the Chapter 2.

2.1.2 S-cone stimuli and SC mediated effects

Several asymmetries have always been reported between S-cones and L- and M-cones (e.g. S-cones are rare, absent from the central fovea, etc.), and S-cone vision have always been described as being "special" (Smithson, 2005).

Importantly for my investigation, electrophysiological studies have reported that the retino-collicular pathway does not contain projections from S-cones, (de Monasterio, 1978; Marrocco & Li, 1977; Schiller & Malpeli, 1977). The S-cones are thought primarily to mediate colour perception, and their signals are carried by morphologically distinct types of ganglion cells, which project to the KC layers of the lateral geniculate nucleus and thence to layers 2 and 3 of the striate cortex (Dacey & Lee, 1994; White, Wilder, Goodchild, Sefton, & Martin, 1998; Hendry & Reid, 2000; Sun et al., 2004).

Since the MC pathway (projecting to the SC) is not colour opponent and receives little input from S-cones (Calkins, 2001; Chatterjee & Callaway, 2002; Gouras, 1968; Stockman et al., 1991) the initial sensory activity in SC is colour blind (Marrocco & Li, 1977; Ottes, Van Gisbergen, & Eggermont, 1987). Chromatic changes visible only to S-cones should be invisible to both retino-collicular and MC pathways (Sumner, 2006; Sumner et al., 2002; Sumner, Nachev, Vora, Husain, & Kennard, 2004).

Therefore, the role of the SC in different tasks could be tested by selectively activating the S-cones. S-cones may still provide some input to the SC and to the luminance pathway, as suggested by some researchers (Calkins, 2001; Stockman et al., 1991). But such input is small and can be masked by luminance noise (Barbur et al., 1994; Birch et al., 1992; Mollon, 1982). Given this assumption, several authors tried to manipulate the chromaticity and luminance of the stimuli, in order to indirectly investigate the role of the SC. Here I report some behavioural studies that adopted the S-cone method and their results.

Corballis (1998) found with split-brain subjects that the redundancy gain was greatly diminished when the stimuli were equiluminant with the background (that is, presumably undetected by the SC). Later, Sumner et al. (2002) applied the S-cone method to investigate two different effects mediated by the SC: the "oculo-motor distractor effect" (the slowing down of saccadic reaction times to a target when a distracting stimulus is

presented simultaneously; Lévy-Schoen, 1969) and the visual capture of attention (studied with the Posner cuing paradigm). The authors used S-cone and luminance stimuli, assuming that the first would not have access to the SC while the second ones would. They found a lack of oculo-motor distractor effect with S-cone stimuli proving the crucial role of this structure in the distracting effect. On the contrary, a pure S-cone chromatic stimulus had a clear influence in the Posner cueing task, showing that the SC is not crucial for the capturing of the exogenous attention (contrary to the models that hypothesise the involvement of the MC pathway in exogenous orienting processes).

Savazzi and Marzi (2004) investigated the involvement of the SC in the redundant target effect in both normal subjects and patients with a total section or agenesis of the corpus callosum, with both luminance and S-cone stimuli. Violations of the Race model inequality test (for a description, see Chapter 1) were found with luminance stimuli both in normal observers and in patients lacking the corpus callosum, evidencing a co-activation model explanation. However, when short-wavelength stimuli were used (purple or mixed white/purple), no violation occurred, and the probability summation could better explain the data. In other words, when detection occurs as a result of activity in the PC or MC pathways, neural co-activation occurs. However, when the KC pathway is activated, without direct input to the SC (Marrocco & Li, 1977), probability summation occurs (Savazzi & Marzi, 2002; 2004).

The evidence on this topic is not conclusive. For example, in a study by Turatto, Mazza, Savazzi, & Marzi (2004) a difference in the redundancy gain for stimuli mediated by the PC and MC pathways was reported. The authors used different stimuli in order to selectively activate the PC and MC pathways: static feature singletons in the first case and visual transients in the second case, claiming that the PC pathway would activate mainly the extrastriate visual cortex (De Weerd et al., 2003), whereas the MC pathway would activate the SC. The results showed that the mechanism underlying the redundancy gain can shift from co-activation (violation of the model) to probabilistic (statistical facilitation) depending upon whether target detection is carried out mainly by the PC and MC visual streams, respectively. The reason of the discrepancy between their data and other data present in the literature is unclear and is probably related to specific stimulus and task

conditions. However, the analysis of the reasons is beyond the scope of the present thesis, in which I will focus mainly on studies that adopted the S-cone method.

In a more recent study Leo, Bertini, di Pellegrino and Ladavas (2008) asked healthy observers to detect a target (either red or purple), with a task-irrelevant auditory stimulus that could have been either spatially coincident or non-coincident. The spatial congruency effect occurred with red but not purple targets; the MRE effect was super-additive only for red targets in the congruent condition, proving that the spatial rule is mediated by the activation of the SC.

In conclusion, it is worth pointing out that all those studies have been conducted within a *random target paradigm*, in which the participants had to divide their attention on both the visual and auditory modalities. Whether the results can be replicated within a *focused attention paradigm* (with the attention being focused only to one modality, while the second being task-irrelevant), is still uncertain and will be discussed in the present chapter.

In the next paragraphs I will first introduce some methodological aspects of the experiments described in the present chapter (e.g. minimum motion technique and tritanopic technique), followed by the description of the experiments and a discussion.

2.2 Equiluminance

Since the main manipulation of the experiments that I will present in the present chapter regards the isolation of the colour pathway (i.e., the use of stimuli that minimise the activation of the luminance pathway) I will report here a brief introduction on the methodology that I adopted.

In general, given a middle grey background it is simple to create stimuli that activate only the luminance pathway (by using stimuli with higher or lower luminance, but the same spectral composition of the background), whereas it is more challenging to create stimuli that activate only the colour pathway without being visible to the luminance pathway. In the literature, a stimulus that, regardless of its chromaticity, produces no

response in the luminance pathway is said to be *equiluminant* (or *isoluminant*) with its background.

Two aspects of equiluminance are relevant to the following experiments: a) different targets will be made equiluminant, one relative to the other, to make sure that they are equally efficient (i.e., that detection latencies do not differ in the two cases); b) targets that should not activate the SC according to the S-cone method will be made equiluminant relative to the background.

Equiluminant stimuli are usually described as blurry, producing only weak motion and depth perception and contributing very little to the perception. An equiluminant image provides the visual system with information about colour, but not about edges and shapes. The point of equiluminance may be different from individual to individual, varies across the retina and varies with the variation of temporal and spatial frequencies (Livingstone & Hubel, 1987; Mullen, 1991). In general, it is important to clarify why two equiluminant stimuli can have different luminance values.

Luminance is a photometric measure that describes the amount of light (i.e., the electromagnetic radiation weighted by the photopic sensitivity curve) that passes through or is emitted from a particular area, and falls within a given solid angle. The SI unit for luminance is the *candela per square metre* (cd/m^2).

The appropriate variable for describing the different luminances of a stimulus and its background is called *contrast*: that is, the variation of the amount of light in a stimulus relative to its background, normalized by the overall amount of light. The amount of contrast is usually calculated by the Michelson contrast formula (Michelson, 1927):

$$\mathbf{MC = (I\ target\ -\ I\ background)/(I\ target\ +\ I\ background)}$$

with MC varying between -1 and 0 for decrements, and between 0 and +1 for increments.

Two regions with the same chromaticity and the same luminance contrast will appear identical. However, two regions with different chromaticities (i.e. two heterochromatic stimuli) and the same luminance contrast may differ not only in their hue

and saturation, but also in their brightness, which is the subjective counterpart of the luminance contrast (Corney, Haynes, Rees & Lotto, 2009). For example, the Helmholtz-Kohlrausch effect (Wyszecki, 1967; Nayatani, 1998) shows that the brightness of a stimulus is not a simple representation of luminance, since the brightness of equiluminant stimuli changes with their relative saturation (i.e., strongly coloured stimuli appear brighter than grey stimuli), and with shifts in the spectral distribution of the stimulus (e.g., blues and reds appear brighter than greens and yellows at equiluminance; Pridmor, 2007).

As regards the use of reaction time (RT) as the main dependent variable in the following experiments, it is well known that RTs are influenced by the luminance contrast. The relation is defined by Pieron's law (Pieron, 1952), which can be expressed as follows:

$$RT = R_0 + kI^{-\beta}$$

where R_0 is the asymptotic *RT*, *I* is the physical intensity of the stimulus (e.g., luminance contrast), *k* and β are constants.

Pieron's law has been used to describe the separate effects of chromatic and luminance contrasts (e.g. Hsu, 2005). However, it is still uncertain how RT depends on their combined effects. As a first approximation, I assumed that heterocromatic stimuli made equiluminant by a conventional method were equally efficient. For this purpose I adopted two different methods. In Experiment 1 equiluminance was based on the *minimally distinct border* method, in which the observer adjusts the luminance of a target region until the border with an adjacent comparison region is minimally salient. In Experiment 2 I used the *minimum motion technique* (Anstis & Cavanagh, 1983), explained in next paragraph.

In the S-cone method, described in the previous paragraph, a stimulus intended to activate only (or mainly) the S-cones should be equiluminant to its background. A nominally equiluminant stimulus (as defined by an independent procedure) may be detectable by the luminance pathway because of a residual luminance difference that can arise for many reasons: erroneous stimulus calibration, variations in individual observer's equiluminant point, or variations in the equiluminant balance of the individual neurons (Dobkins & Albright, 1994; Gegenfurtner et al., 1994; Logothetis & Charles, 1990). In order to mask the residual luminance, I will use two different techniques: the *luminance noise*

(Barbur et al., 1994; Birch et al., 1992; Mollon, 1982) in Experiments 1 and 2, and the *tritanotopic technique* (Cavanagh et al., 1987) in Experiment 3.

2.2.1 The minimum motion technique

To find the point of equiluminance between two different colour stimuli Anstis and Cavanagh (1983) introduced the *minimum motion technique*, which is based on a 4-frame animation in which two frames include a grating with a visible luminance contrast and the two alternating frames include a grating with an adjustable luminance contrast between chromatically different stripes, as shown in Figure 2.1.

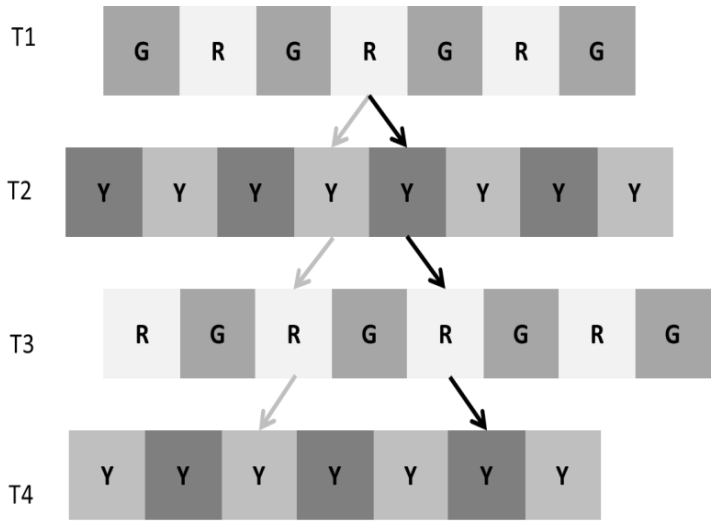


Figure 2.1 The minimum motion technique. Four coloured gratings are shown in a repetitive sequence (from T1 to T4 in a loop), each displaced for half of each square width from the preceding one. The sign of the luminance contrast between the two target colours (green G and red R in this case) determines the direction of the apparent motion (leftward or rightward), allowing to find the subjective point of equiluminance in the case of no motion. A) Apparent motion occurs in the leftward direction when the red square, being lighter that the green one, jumps towards the light yellow, while the green square, being darker that the red one, jumps towards the dark yellow. B) Apparent motion occurs in the rightward direction when the red square, being darker that the green one, jumps towards the dark yellow (Anstis & Cavanagh, 1987).

An apparent motion, toward the left or toward the right side, is seen when a luminance difference is present between the two target colours (e.g., red and green).

Motion direction depends on the sign of the luminance contrast: if the red is darker of the green, perceived motion will be consistent with the unification of the red with the dark yellow of the subsequent grating, and vice versa. When the target colours have the same luminance no motion is perceived.

2.2.2 Random luminance technique

The principle underlying the *random luminance technique* (or also called *random luminance noise*; Barbur, Harlow, & Plant, 1994; Birch, Barbur, & Harlow, 1992; Mollon, 1982) is simple and intuitive, and many other studies have adopted this technique for the masking of residual luminance contrast and the improvement of S-cone isolating stimuli (Leo et al., 2008). To isolate the chromatic component of a blue target (Figure 2.2) and nullify every residual luminance contrast with the grey background, one can embed it in a matrix of flickering grey cells that change randomly their luminance contrast over time. Such a flickering luminance noise will make the threshold for detecting luminance contrast in the luminance channel higher. Any residual luminance contrast produced by the blue stimulus will be below threshold, and the visual system will detect the target only because the colour contrast activated the colour pathway.

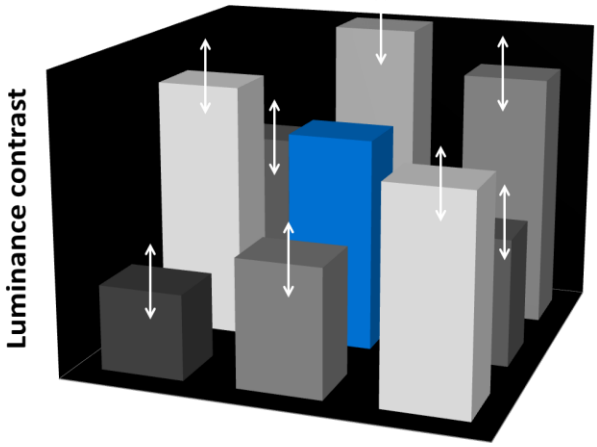


Figure 2.2 Random luminance technique. The luminance contrasts of the 8 cells surrounding the blue target independently change over time.

2.2.3 Tritanopic technique

The tritanopic technique is an alternative method aiming at the isolation of S-cone colour information from luminance information. The word *tritanope* refers to an individual who lacks S-cones. The stimuli invisible to a tritanope are therefore called tritanopic. According to this method, a blue stimulus displayed on a yellow adaptive background should activate the chromatic pathway without activating the luminance pathway (Figure 2.3).

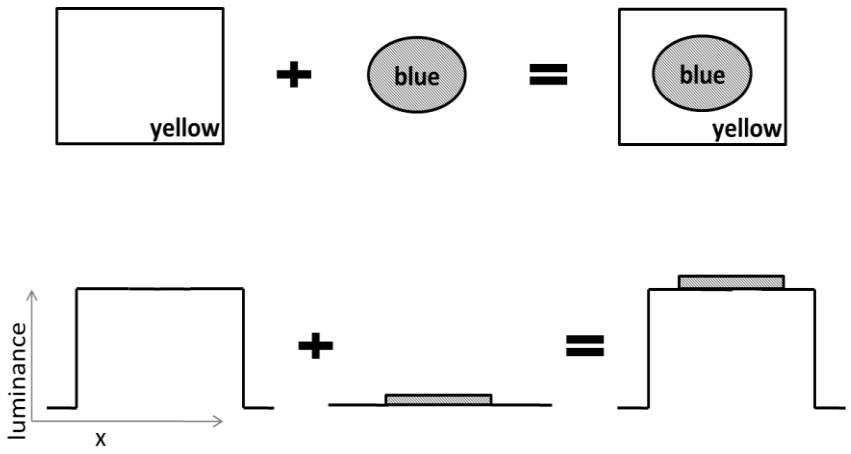


Figure 2.3 Illustrative representation of a tritanopic stimulus. From *Cavanagh, Adelson, & Heard (1992).* A tritanopic stimulus is created presenting a blue stimulus on a yellow adapting background. Even if there is a luminance variation created by the presentation of the blue image, it is below the high luminance threshold created by the yellow adapting field. In this way, only S-cones respond to the stimulus, and the image in nearly- equiluminous.

The yellow field drives the response of M and L cones to a high level, decreasing the sensitivities of these channels, and thereby permitting the short wavelength mechanism's sensitivity to be isolated (Figure 2.4). Since the M and L cones constitute the principal source to the luminance channel, such an image will be effectively equiluminant, even if the blue stimulus will be perceived as a bright field.

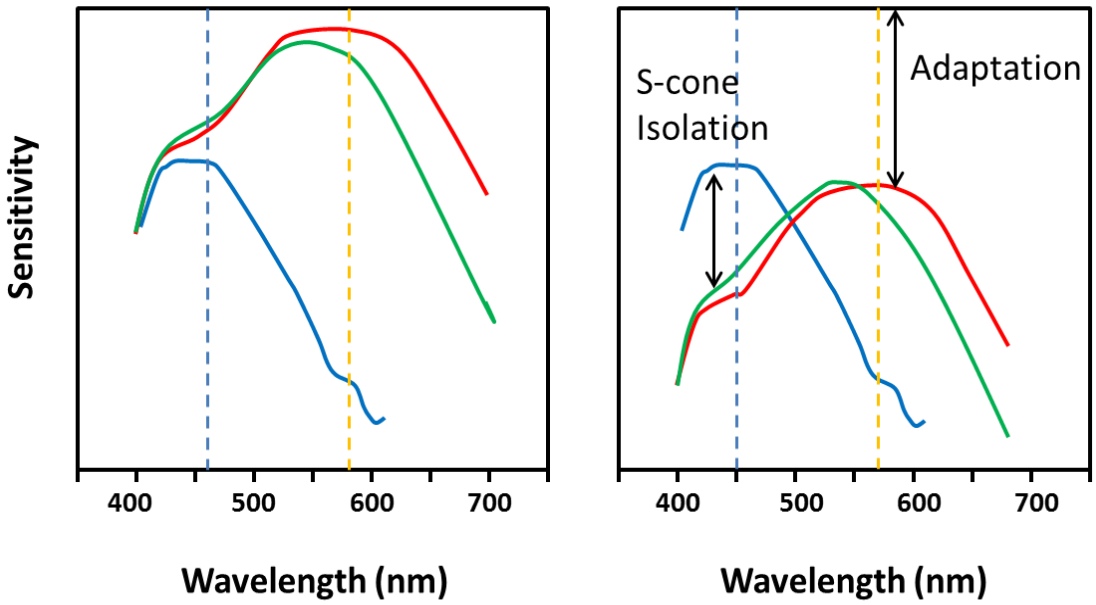


Figure 2.4 Spectral sensitivities and tritanopic stimuli. The left graph shows the spectral sensitivities of the three cones (S- „blue“, M -„green“ and L -„red“) under normal conditions. The vertical blue line indicates the peak wavelength of the S-cone stimulus and the vertical yellow line indicates the peak wavelength of the background. Under these conditions, sensitivity to a short wavelength stimulus is higher for M- and L-cones than for S-cones. The right graph shows the spectral sensitivities of the three cones after adaptation to a bright broadband yellow background. As a consequence of the decreased sensitivities of M- and L-cones, the sensitivity to a short wavelength stimulus is higher for S-cones.

2.3 The spatial rule in the audio-visual integration mediated by the SC

2.3.1 Rationale of the experiments

Three experiments were run with the aim of investigating the spatial rule of multisensory integration in a focused attention paradigm in which participants should explicitly focus on visual targets, ignoring the auditory modality. To achieve this, the S-cone method (Sumner et al., 2002) was applied to the stimuli in order to create SC- effective (red) and SC-ineffective (blue) visual targets. All trials contained visual patterns displayed at the left/right of the central fixation point, marking the possible location of the target and favouring the distribution of spatial attention over a broad area. In positive trials such patterns were followed by a red/blue target (with/without a task-irrelevant lateralized auditory stimulus, in multisensory/unisensory trials) while in negative trials they were followed by a task-irrelevant lateralized auditory stimulus. The peculiarity of such a task is that the participants, besides ignoring the auditory stimulus, had also to inhibit the response to it in the negative catch trials (in which only the auditory stimulus were present). The aim of using this task was to investigate whether MI is an automatic process, and wheter an auditory stimulus can influence the detection of a visual stimulus even when it is task-irrelevant (and when the participant is explicitely required to refrain any response toward the auditory stimulus alone). Moreover, I expected that any influence of the auditory stimulus would appear only when the visual stimulus will be detected by the SC (thus, the red target) and when the unattended auditory stimulus will be spatially congruent with the visual stimulus (following the spatial rule of the multisensory integration).

2.3.2 Experiment 1

The experiment was based on the comparison of go responses in three sets of positive trials:

- Unisensory visual (US): the visual (red or blue) target was presented alone, on the left/right of the fixation cross;
- Multisensory same position (MS_{sp}): the visual (red or blue) target and the auditory stimulus were presented simultaneously in the same spatial position, on the left/right of the fixation cross;
- Multisensory different position (MS_{dp}): the visual (red or blue) target and the auditory stimulus were presented simultaneously but in opposite positions (A-left and V-right, or vice versa).

To make responses contingent on the presentation of visual targets, the following set of catch trials were added, in which no response was required:

- Unisensory auditory: the auditory stimulus was presented alone, from either the left or right loudspeaker.

I expected participants to be equally accurate in all experimental conditions, if targets of different colours were matched for efficiency. However, a combination of effects are expected on RTs: no difference between red vs. blue targets in US trials; a significant MRE effect in both MS_{sp} and MS_{dp} conditions; a larger MRE effect for red targets in the MS_{sp} condition; no difference between red vs. blue targets in MS_{dp} trials.

2.3.2.1 Method

Participants

Fourteen healthy right-handed female observers (age range 19-25 years) took part in the experiment. All had normal hearing and normal or corrected-to-normal visual acuity. Participants were students, naïve as to the purpose of the experiment, who received

course credit for their participation. They gave their informed consent prior to the beginning of the experiment.

Stimuli and procedure

Auditory and visual stimuli were generated by a PC equipped with standard acoustic and graphic software. SUPERLAB 4.8 was utilized to present stimuli and collect responses. Visual stimuli were displayed on Sony Trinitron CPD-G200P CRT 17 inch monitor, set at 50% brightness, 90% contrast, 1280 $\times$ 768 pixel resolution, 60 Hz. CIE1931 coordinates of monitor primaries were as follows: R (x= 0.623, y= 0.339); G (x= 0.284, y= 0.589); B (x= 0.152, y= 0.067). The participant was seated at a distance of 57 cm from the centre of the screen and required to hold constant fixation on a central white cross (70 cd/m²) against a black background (5 cd/m²). The task consisted in pressing the left button of the mouse with the index finger of the dominant hand as quickly as possible to respond to any visual target (either red or blue, in either unisensory or multisensory conditions) briefly shown at the left/right of the fixation cross with a 11.5° eccentricity, and suppressing any overt response to the auditory stimulus alone. Target stimuli were 1° $\times$ 1° red (long-wavelength) or blue (short-wavelength) squares, displayed for 67 ms as the central cell of a 3° $\times$ 3° matrix. Each cell of the matrix subtended a 1° $\times$ 1° visual angle. Figure 2.5 shows the sequence of events in a trial: the fixation point lasted 500 ms; after an interval with a random duration of 133-1132 ms, the two matrices were displayed for 733 -1000 ms, refreshing the luminances of the 8 peripheral cells every 67 ms by randomly picking up a value in the 6.4 -10.8 cd/m^2 range, with the only exception of the last 67 ms, in which the 8 luminances of the peripheral cells were fixed (average luminance= 8.6 cd/m²); then, the target stimulus was displayed in the central cell. This random luminance modulation technique allowed us to isolate the chromatic contribution of blue targets from their associated luminance component (Mollon, 1982; Birch et al., 1992, Barbur, 2004).

To generate red/blue stimuli with the same efficiency I adopted the *minimally distinct border technique* (Kaiser, 1971) in which a variable comparison hemidisk was matched to an adjacent constant hemidisk. The diameter of the disk subtended 3°. Each

participant adjusted a grey comparison hemidisk in the 4.4 - 20.7 cd/m^2 range to match it to a 8.6 cd/m^2 blue hemidisk and then a red comparison hemidisk in the 5.5 - 15.3 cd/m^2 range to match it to a grey hemidisk set at the luminance obtained in the first match. In this way the contrast of red and blue stimuli (relative to the average luminance of the surrounding grey cells) was made subjectively equal for each participant. Following this procedure, the luminance of red targets ranged between 5.9 and 14.9 cd/m^2. The auditory stimulus was a 4 kHz pure-tone burst of 60 dB emitted for 67 ms by one of the two piezoelectric loudspeakers (4 W) located 5 cm above the positions of the matrices, where the visual target might appear.

After 20 min of dark adaptation the participant entered a 38-trial training session in which three criteria should be reached: a maximum of two RTs over 500 ms, a maximum of two RTs under 120 ms, and a maximum of two errors (either misses or false alarms). If the participant failed to meet any of the three criteria, the training session was repeated. The experimental session included 416 trials (104 for each of the 4 conditions).

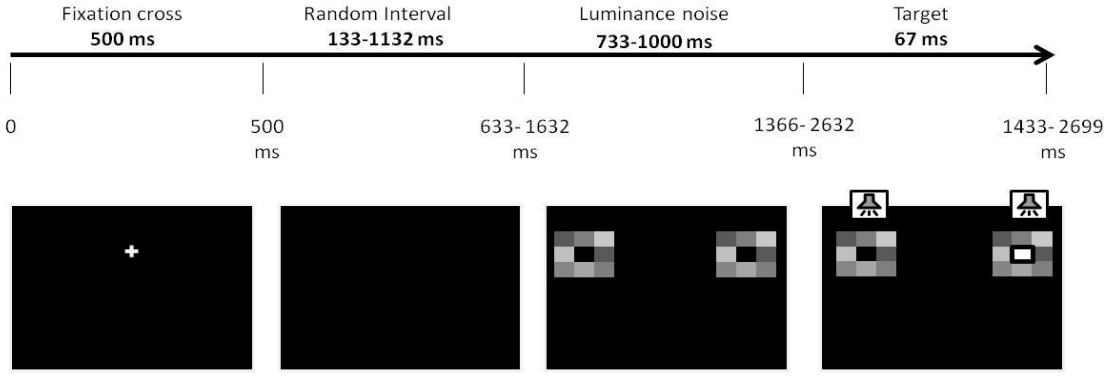


Figure 2.5 Experiment 1, stimuli and timetable of a typical trial. Schematic timetable of a positive trial in Experiment 1 (not in scale). Each trial was initiated by the onset of a central fixation cross. The visual target was shown after an interval varying between 1367 and 2633 ms, including a blank field and the display of two grey-cell matrices. The target was a red or blue square, corresponding to the central cell of one of two matrices. Loudspeakers were placed just above the matrices. In multisensory positive trials an auditory stimulus was presented simultaneously with the visual target, in the same or opposite position. Negative (catch) trials were characterized by the absence of the visual target and presence of the auditory stimulus.

2.3.2.2 Results

Figure 2.6 illustrates the pattern of Misses percentage (lower graph) and RTs (corresponding only to the Hits of the participant; upper graph) in the 6 conditions of the Stimulus Condition × Colour design. The average False Alarm percentage in catch trials was 1.3%. A preliminary test of the distribution of Misses against zero, run on transformed data [x= arcsin(pMiss)$^{0.5}$], showed a significant difference in 5 out of 6 cases (t values larger than 3.81, df= 13, one-tailed, p smaller than the 0.008, critical value chosen after the Bonferroni correction). The only Miss percentage not significantly different from zero was $MissUS_{blue}$= 2.89% (t(13)= 2.16, p< 0.05). A 2-way ANOVA on transformed data showed no main effects of Condition (F< 1) and Colour ($F_{1,13}$= 1.17, p= 0.3), but a significant 2-way interaction ($F_{2,26}$= 4.28, p< 0.05), attributable to worse detection of red than blue targets in unisensory but not multisensory trials ($MissUS_{red}$ vs. $_{blue}$= 6.18 vs. 2.89%: t= 4.17, p< 0.01; $MissMS_{sp\_red}$ vs. $_{sp\_blue}$= 5.63 vs. 3.99%: t(13)= 0.48, p= 0.7; $MissMS_{dp\_red}$ vs. $_{dp\_blue}$ = 4.26 vs. 4.53%: t(13)= 0.56, p= 0.6; two-tailed, for all post hoc comparisons). As regards accuracy, the unexpected superiority of blue targets could be attributed to an inadequacy of the procedure for matching stimulus effectiveness.

As to RTs, all ANOVAs were performed on speed values of Hits (speed= 1/RT), although RTs in ms will be reported in the text and figures to facilitate the comparison of mean values in different conditions. Responses in US trials were faster for blue than red targets (RT_{US_blue}= 340 ms vs. RT_{US_red}= 358 ms: t(13)= 3.63, two-tailed, p< 0.01). This result parallels the blue-target superiority found in accuracy and suggests that the minimally distinct border method failed to match stimulus saliency. Taking reaction time in US trials as a baseline, I concluded that blue targets were more efficient than red targets.

To evaluate the amount of MRE I ran a 3 × 2 within-subjects ANOVA with Condition (US, MS_{sp}, MS_{dp}) and Colour (red, blue) as factors. The main effect of Condition ($F_{2,26}$= 71.50, p< 0.01) was dependent on the presence of a significant MRE effect in both MS conditions (RT_{US}= 349 ms vs. RT_{MS}= 318 ms: $F_{1,26}$= 142.89, p< 0.01); mean RTs in the two multisensory conditions did not differ (F< 1). I found no main effect of Colour ($F_{1,13}$= 1.27, p= 0.2). However, the Condition × Colour interaction was significant ($F_{2,26}$= 11.13, p< 0.01)

and could be attributed to the presence of a blue over red superiority in US, but not MS, trials (RT_{MS_red}= 317 ms vs. RT_{MS_blue}= 320 ms: t(27)= 1.37, two-tailed, p= 0.2).

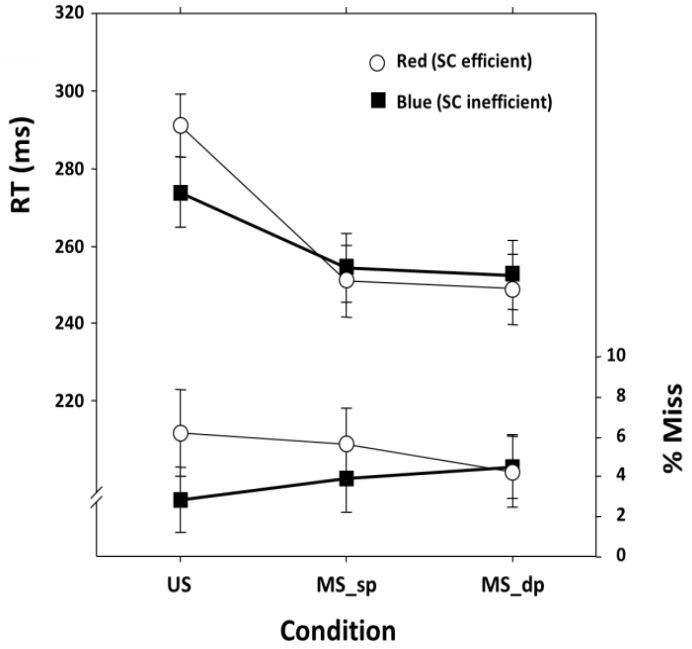


Figure 2.6 Experiment 1, % Miss and RT. Mean Miss percentages and RTs (± 1 s.e.m.) in the three sets of positive trials of Experiment 1, for red and blue targets. In US trials responses to blue targets were faster and more accurate; in general, responses were faster in MS than US trials.

To confirm such conclusions I analysed the amount of MRE as a function of target colour and spatial congruency of the acoustic source, using the **%MRE index** (Stein & Meredith, 1993), adapted for the focused attention task:

$$\frac{(\mathbf{US - MS})}{\mathbf{US}} \times \mathbf{100}$$

Figure 2.7 shows the pattern of %MRE values for each condition of the 2 × 2 designs with Spatial position (sp, dp) and Colour (red, blue) as factors. Average %MRE values were always positive, suggesting that multisensory stimulation was always associated with a significant speed gain in both spatial conditions, for both red and blue targets (all four t values larger than 4, df= 13, one-tailed, p< 0.01). A 2-way ANOVA on %MRE values showed that the MRE was significantly larger for red than blue targets ($\%MRE_{red}$= 6.2% vs.

%MRE_{blue}= 3.1%: $F_{1,13}$= 13.42, p< 0.01). The spatial position of the acoustic source had no effect: neither the main effect of Spatial position ($F_{1,13}$= 2.30, p= 0.15) nor the Spatial position × Colour interaction (F< 1) were significant.

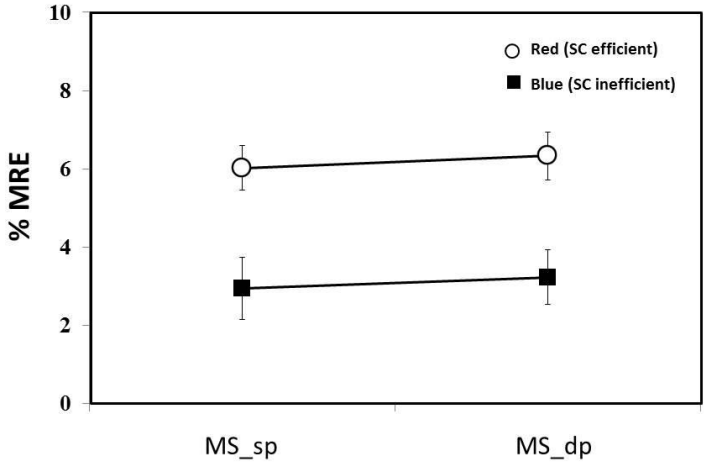


Figure 2.7 Experiment 1, %MRE (± 1 s.e.m.). The distribution of %MRE indicates that the %MRE was stronger for red that for blue stimuli. No effect of the spatial position of the sound is evident.

As regards the pattern of RTs, the expectation about the multisensory enhancement was fulfilled: responses in both MS_{sp} and MS_{dp} trials were faster than in US trials. Unexpectedly, a superiority of blue over red targets was found in US trials, consistent with accuracy data and the hypothesis that the procedure for matching the target effectiveness introduced a systematic bias resulting in a weaker red target. As regards the multisensory enhancement, I found a significant effect for both targets and a clear superiority of red targets, but no difference as a function of spatial congruency.

2.3.2.3 Conclusion

Results suggest that when the task is performed in a focused attention task, the MRE does not follow the rules of the MI mediated by the SC. However, before drawing any definitive conclusion, some methodological imperfections will be pointed put. First of all,

the amount of MRE was independent of the position of the auditory stimulus, possibly because of an insufficient distance between the sources of auditory and visual stimulations in incongruent trials. Moreover, a posteriori I have realised that there may be some critical weakness in the method used for the creation of "SC-effective" (red) and "SC-ineffective" (blue) stimuli. In fact, red and blue stimuli may not have been functionally different in terms of their likelihood of activating the SC. Since both the red and blue stimuli were matched for brightness with the grey background (with the minimally distinct border method) and in luminance noise, they might be both visible only to the chromatic channel. Since there is no clear evidence that any of the direct pathways to the SC are chromatic, they may have been similar in term of effectiveness in the SC. The SC, however, appears to distinguish colour only after a delay, suggesting that target information based on colour is elaborated elsewhere first. However, using clear luminance stimuli (e.g. white stimuli), one may create a strong discrepancy in terms on efficiency in comparison to pure S-cone colour stimuli. This is why my aim was to minimise the difference between the two targets, still making possible for one to selectively activate the SC while the second would avoid the SC.

2.3.3 Experiment 2

Taking into account the methodological weaknesses in Experiment 1, three differences have been introduced in Experiment 2: first, a larger eccentricity of the stimuli were introduced (15.5° instead of the 11.5° from Experiment 1); second, the fixation cross was moved 5° below the horizontal line connecting the lower border of the two matrices, to prevent visual targets from falling in the blind spot of the ipsilateral eye; third, red and blue targets were matched for their brightness using the *minimum motion technique* (Anstis & Cavanagh, 1983). Moreover, it is known that the point of equal brightness changes with the eccentricity. In Experiment 1 this has not been taken into account: the point of equal brightness was measured in central vision, while the stimuli during the experiment were presents peripherally. This methodological imperfection might explain the absence of the main effect. In Experiment 2, the equal brightness was measured at peripherally, separately for left and right side.

Importantly for the purpose of the experiment, the red and blue targets were not matched with the grey background of the luminance noise. In this way, even if the red and blue were of equal brightness among them, the luminance contrast with the grey background was not the same. I assumed that the input to the luminance channel produced by the S-cone stimulus (blue) would be weak and would be masked by the luminance noise, while the luminance contrast of the ML-cone stimulus (red) will be only partially masked by the noise, and part of it will still stimulate the luminance pathway.

2.3.3.1 Method

Participants

Eleven observers participated in Experiment 2. Nine had already participated in Experiment 1, while the other two were new to the task (20 and 24 year old, both right-handed females), also with normal hearing and normal or corrected-to-normal visual acuity.

Stimuli and procedure

Except for eccentricity, stimuli were like in Experiment 1. The matching of red and blue target luminance was achieved using the *minimum motion technique* (instead of the minimally distinct border technique of Experiment 1). Four frames were exposed in a repetitive sequence: frames 1 and 3 contained opposite-phase luminance gratings composed of light/dark green squares (18.5 and 5.5 cd/m^2), while frames 2 and 4 contained gratings composed of red/blue squares with a quarter-cycle phase shift. When red and blue luminance differed, motion was perceived towards either the left or right according to spatiotemporal proximity between homologous border contrasts. The luminance of blue squares was set to 8.6 cd/m^2.

The individual equiluminance point to be utilized in red-target trials was computed as the average of the adjustments obtained in a preliminary session in which the participant asked the experimenter to adjust the brightness of red squares in comparison

to blue squares until sideways motion was substituted by flicker. The luminance of red targets resulting from the application of the minimum motion technique ranged between 7.7 and 12.0 cd/m^2. I can assume that a luminance signal was sent to the SC for red stimuli (given by the luminance contrast, either positive or negative, with the background), while the blue stimuli were perfectly equiluminant with the background. All other aspects of the procedure and data analysis were identical to those in Experiment 1.

2.3.3.2 Results

Figure 2.8 illustrates the pattern of Miss percentage and RTs of Hits in the 6 conditions of the Condition × Colour design. The average False Alarm percentage in catch trials was 1.9%. A preliminary test of the distribution of Misses, run on transformed data [x= arcsin(pMiss)$^{0.5}$], showed that no mean Miss percentage differed from zero (all t values smaller than 2.52, df= 10, one-tailed, all p values larger than the 0.008 critical value chosen after the Bonferroni correction). Therefore, subsequent analyses took into account only RTs.

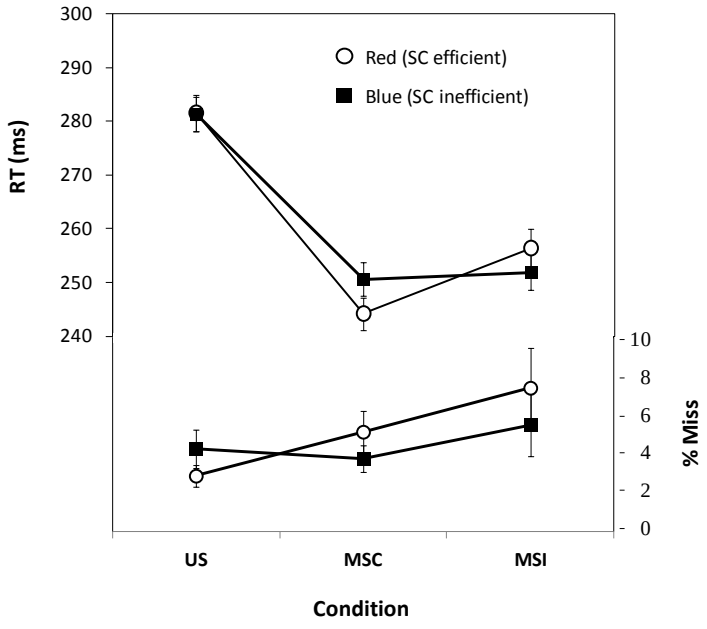


Figure 2.8 Experiment 2, % Miss and RT. Mean Miss percentages and RTs (± 1 s.e.m.) in the three sets of positive trials of Experiment 2, for red and blue targets. In US trials the target colour did not affect RTs and errors. RTs were faster in MS than US trials, while the number of errors increased significantly in the multisensory conditions for red but not blue targets.

In US trials no significant difference was found between response speed for red and blue targets (RT_{US_red} vs. $_{US\_blue}$= 348 vs. 348 ms: t(10)= 0.13). The 3 × 2 within-subjects ANOVA showed a main effect of Condition ($F_{2,20}$= 48.23, p< 0.01), no main effect of Colour (F < 1), and of the Condition × Colour interaction ($F_{2,20}$= 2.53, p= 0.1). Nevethless, on the basis of the strong predictions I had about the data, I performed a planned-comparison analysis, showing that the spatial congruency of the acoustic source affected the detection of red ($RT_{MS\_sp\_red}$ vs. $_{MS\_dp\_red}$= 311 vs. 323 ms: $F_{2,20}$= 10.06, p< 0.01) but not blue ($RT_{MS\_sp\_blue}$ vs. $_{MS\_dp\_blue}$= 317 vs. 319 ms: F< 1) targets.

Figure 2.9 shows the pattern of MRE amounts, expressed as %MRE, in the four conditions of the Spatial position × Colour design. As in Experiment 1, all four average %MRE values were positive, confirming that multisensory stimulation always produced a speed gain, for both red and blue targets and in both multisensory congruency conditions (all t values larger than 5, df= 10, one-tailed, p< 0.01). A planned-comparison analysis of the four average %MRE percentages confirmed that MRE was stronger in the sampe position (SP) than different position (DP) condition for red (5.7 vs 3.8%: $F_{1,10}$= 8.02, p< 0.02) but not blue (4.6 vs 4.5%: F< 1) targets, consistently with the results of the 2 × 2 ANOVA approaching a level of significance (Spatial position: $F_{1,10}$= 4.14, p= 0.07; Colour: F< 1; Spatial position × Colour: $F_{1,10}$= 3.57, p= 0.09).

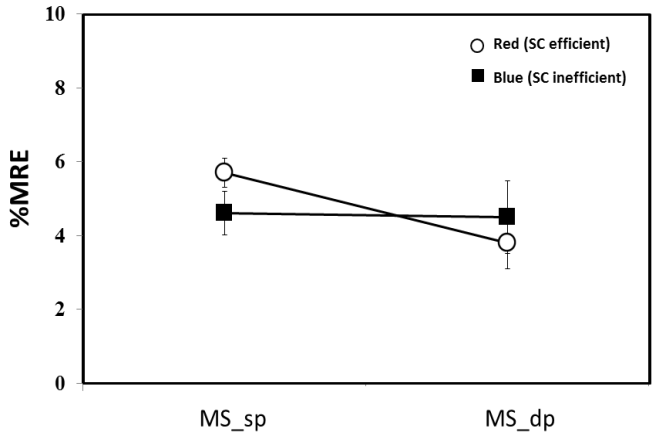


Figure 2.9 Experiment 2, % MRE (± 1 s.e.m.). The distribution of %MRE indicates that performance was better in MS_{sp} than MS_{dp} trials with red but not blue targets

The RT pattern followed initial hypothesis. In US trials responses to red and blue targets were equally fast. A multisensory enhancement was obtained in both MS_{sp} and MS_{dp} conditions, and it was larger in MS_{sp} than MS_{dp} trials only for red targets. Furthermore, in MS_{sp} trials responses were faster for red than blue targets, as expected if the spatial rule applies and multisensory integration is mediated by the SC.

2.3.3 Experiment 3

To decrease the responsiveness of the luminance pathway and maximize the differential efficiency of the two coloured targets, in this experiment I adopted the *tritanopic technique* (Cavanagh et al., 1987) instead of the random modulation technique of Experiments 1 and 2. The intense yellow adapting field shown before a blue increment (Figure 2.10) can isolate the response of the S-cones very effectively (Stiles, 1959; Wald, 1966), as already described in the introductive paragraphs of the preset chapter (paragraph 2.2.3).

Apart from the elimination of the randomly modulated grey-cell matrices and the introduction of the yellow field, Experiment 3 replicated Experiment 2 and was conducted having the same expectations. Experiment 3 should provide converging evidence on the role of the SC, with methods different from those used in Experiment 2.

2.3.3.1 Methods

Participants

Twelve observers participated in Experiment 3. Nine had already participated in both Experiments 1 and 2, while the other three were new to the task (20, 21 and 24 year old, all right-handed, one male and two females), also with normal hearing and normal or corrected-to-normal visual acuity.

Stimuli and procedure

With respect to Experiment 2 there were three main changes in the visual stimuli. The background was yellow (RGB values: 200, 160, 0; 47.1 cd/m2).

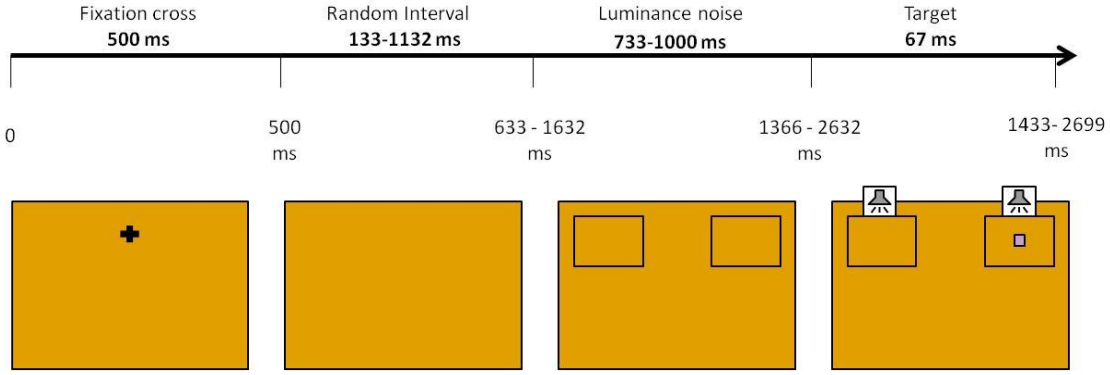


Figure 2.10 Experiment 3, stimuli and timetable of a typical trial. Schematic timetable of a positive trial in Experiment 3 (not in scale). Each trial started with a central fixation cross (500 ms), followed by a variable interval lasting fromm 133- 1132 ms and two outline squares placed on the left and right corner on the top of the screen, with a variable duration ranging from 733-1000 ms. The target was a red or blue square on the yellow background, corresponding to the central cell of one of two matrices, and lasting for 67 ms. Loudspeakers were placed just above the matrices. In multisensory positive trials an auditory stimulus was presented simultaneously with the visual target, in the same or opposite position. Negative (catch) trials were characterized by the absence of the visual target and presence of the auditory stimulus. The tritanopic technique was applied as described in the main introduction.

The grey-cell matrices were substituted by black outline squares (side= 3°). RGB values and luminances of red and blue targets were as follows: red (255, 160, 0; 55.5 cd/m^2); blue (200, 160, 200; 55.7 cd/m^2). The temporal sequence of events included in every trial was the same as in previous experiments. The black fixation cross (5.2 cd/m^2) was displayed in the same position of Experiment 2.

2.3.3.2 Results

Figure 2.11 illustrates the pattern of Miss percentages and RTs of Hits in the 6 conditions of the Condition × Colour design. The average False Alarm percentage in catch trials was 1.8 %. A preliminary test of the distribution of Misses, run on transformed data

[x= arcsin(pMiss)$^{0.5}$], showed that 4 out of 6 mean Miss percentages differed from zero (t values larger than 3.28, df= 11, one-tailed, p values smaller than the 0.008 critical value chosen after the Bonferroni correction). The two mean Miss percentages not significantly different from zero were $MissMS_{\_sp\_blue}$= 1.61% and $MissMS_{\_dp\_blue}$= 1.60% (t(13)= 2.49 and 3.13, *p*< 0.05 and 0.01, respectively). A 2-way ANOVA on transformed data showed that neither the main effect of Stimulation ($F_{2,22}$= 1.50, *p*= 0.24) nor the 2-way interaction (F< 1) were influential, whether Colour had a significant effect ($F_{1,11}$= 12.93, *p*< 0.01), attributable to a larger number of errors for the red target in all conditions [$MissUS_{\_red}$ vs. $_{blue}$= 10.26 vs. 2.73%: t(11)= 2.37, p< 0.05; $MissMS_{\_sp\_red}$ vs. $_{blue}$= 9.45 vs. 1.60%: t(11)= 3.92, *p*< 0.01; $MissMS_{\_dp\_red}$ vs. $_{blue}$= 7.85 vs. 1.60%: t(11)= 2.76, *p*< 0.05;, two-tailed).

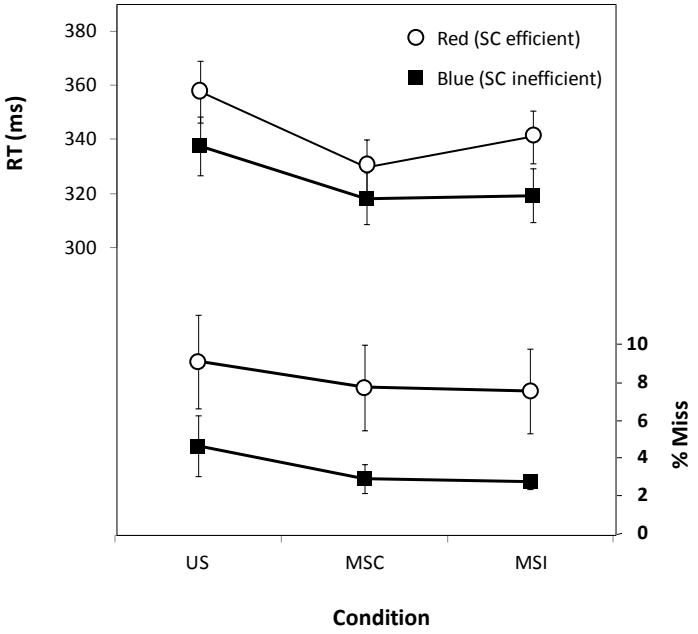


Figure 2.11 Experiment 3, % Miss and RT. Mean Miss percentages and RTs (± 1 s.e.m.) in the three sets of positive trials of Experiment 3, for red and blue targets. In US trials red targets led to more errors and longer RTs. The number of errors was the same in the three positive conditions, and was larger for red than blue targets. RTs were shorter in MS than US trials.

The distribution of RTs was consistent with accuracy data. In US trials go responses were significantly slower for red than blue targets (RT_{US} red vs. blue= 358 vs. 338 ms: t= 7.26, df= 11, two-tailed, *p*< 0.01). The 3 × 2 within-subjects ANOVA showed a main effect

of Condition ($F_{2,22}$= 38.18, p< 0.01) and a main effect of Colour ($F_{1,11}$= 77.44, p< 0.01). The Condition × Colour interaction ($F_{2,22}$= 2.22, p= 0.1) did not reach the significance.

Figure 2.12 shows the amounts of MRE, expressed as Michelson contrast percentage, in the four conditions of the Spatial position × Colour design. As in Experiments 1 and 2, all average %MREs value were positive, indicating that the multisensory stimulation always produced a speed gain, for both red and blue targets, in both congruency conditions (all t values larger than 5, df= 11, one-tailed, p< 0.01). A planned-comparison analysis of the four average %MRE percentages confirmed that the MRE was larger in the SP than incongruent DP for red (4.0 vs. 2.3%: $F_{1,11}$= 11.60, p< 0.01) but not blue (2.3 vs. 2.8%: F< 1) targets. Moreover, in SP trials the %MRE was larger for red than blue targets ($F_{1,11}$= 4.5, p< 0.05).

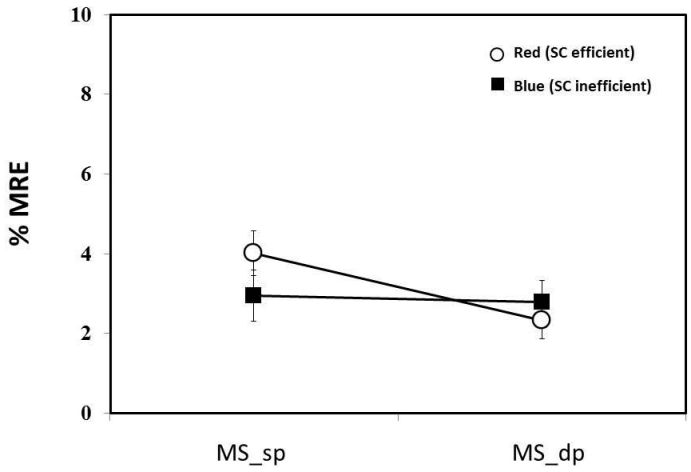


Figure 2.12 Experiment 3, % MRE. The figure shows the distribution of %MRE (± 1 s.e.m.) indicating that, as in Experiment 2, performance was better in MS_{sp} than MS_{dp} trials with red but not blue targets.

Like in Experiment 2 the pattern of MRE values was consistent with the initial hypothesis. The MRE effect was obtained in both MS_{sp} and MS_{dp} conditions; furthermore, its amount was larger in MS_{sp} than MS_{dp} trials only for red targets. Again, in MS_{sp} trials responses were faster for red than blue targets, as expected on the basis of the role of SC in multisensory integration.

2.3.3.3 Comparison of Experiments 1-3

To evaluate the general response pattern over the three experiments and to test a possible effect of practice, data from the 9 observers who participated in all experiments were reanalysed. The distributions of Misses and RTs do not support the hypothesis of a generic learning effect. Performance did not improve over the three experiments. However, there was a slight improvement in Experiment 2 relative to Experiment 1, both in accuracy and response speed. The lack of a further improvement in Experiment 3 might be attributed to large differences in visual stimulus appearance. To support such conclusions I ran two 3 × 3 × 2 within-subjects ANOVA, with Experiment (1, 2, 3), Condition (US, MS_{sp}, MS_{dp}), and Colour (red, blue) as main factors, on $\arcsin(pMiss)^{0.5}$ and 1/RT.

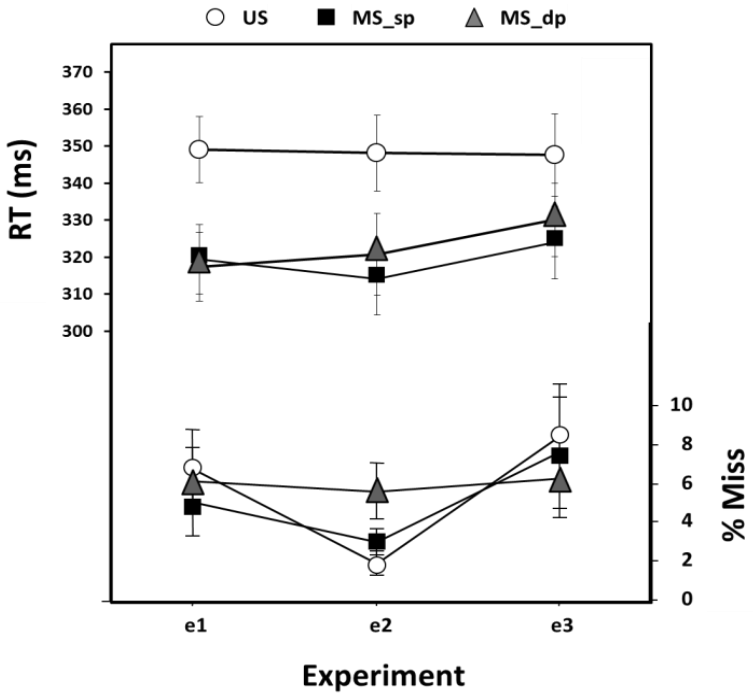


Figure 2.13 Comparison of Experiments 1-3 for RT and %Miss. The RT (ms, upper panel) and % of misses (lower panel; ± 1 s.e.m.) were plotted for the three experimental conditions (US, MS_sp, MS_dp) to evaluate a possible proactive effect over the three experiments.

As regards accuracy, practice had no effect (main effect of Experiment: F < 1). Rather, I confirmed the relevance of target colour. Both the main effect of Colour ($Miss_{red}$ vs. $_{blue}$= 7.41 vs. 3.67%: $F_{1,8}$= 7.76, p< 0.05) and the Experiment × Colour interaction ($F_{2,16}$= 5.43, p<

0.05) were significant. Participants were more accurate in detecting blue over red targets in all experiments; but this superiority did not reach significance in the first two experiments (t(26)= 1.12 and 0.80, corresponding to p= 0.27 and 0.43 for Experiments 1 and 2, respectively; two-tailed), while it was very large in Experiment 3 ($Miss_{red}$ vs. $_{blue}$= 11.77 vs. 2.81%: t(26)= 4.81, p< 0.001), for previously discussed reasons.

As regards RTs, both the main effect of Condition ($F_{2,16}$= 91.82, p< 0.001) and the Condition × Experiment interaction ($F_{4,32}$= 3.24, p< 0.05) were significant, consistently with the superiority of multisensory over unisensory conditions in all experiments and the presence of the spatial congruency effect in Experiments 2 and 3 but not in Experiment 1. There was no main effect of Experiment (F< 1), contrary to a possible practice effect. Both Colour × Experiment ($F_{2,16}$= 13.76, p< 0.001) and Condition × Colour × Experiment interaction ($F_{4,32}$= 4.39, p< 0.05) were significant.

2.4 General conclusion

To provide converging evidence on the role of the SC in multisensory integration I exploited the S-cone method (Sumner et al., 2002) in order to create SC- effective (red) and SC-ineffective (blue) visual targets. In fact, it is widely accepted in the literature that S-cones does not send afferents to the SC. In three experiments two different procedure were adopted for isolating the chromatic component of SC-ineffective (blue) targets from their luminance: the random modulation technique in Experiments 1 and 2, and the tritanopic technique in Experiment 3. Even if this method has been already largely adopted by other authors, my experiments aimed to give additional evidence on the role of the SC in a *focused attention task*, in which the attention is oriented only toward the visual modality.

In Experiment 1 the pattern of %MRE displayed only a superiority of the red over blue condition, but no effect of the spatial disparity. This may be due to the fact that the horizontal separation between AV stimuli in incongruent trials was not large enough to prevent multisensory integration. However, this idea, although plausible, is not consistent with other studies. For example, Leo et al. (2008) found a spatial congruency effect with SC-effective (red) stimuli with an eccentricity of 12° (similar to the eccentricity used in Experiment 1. Given the high similarity between the stimuli adopted in my experiment and those adopted by Leo et al. (2008), the reason for such discrepancy could be attributed to the difference in the paradigm adopted in the two studies. As stated in the introduction, a focused attention paradigm was adopted in my experiments (participants should explicitly focus on the visual modality and detect relevant targets, while ignoring the auditory modality), while other authors (e.g. Leo et al., 2008) adopted a redundant target paradigm, in which the attention is distributed to both modalities. Such a condition, in which the subject is required to focus and respond also to the auditory stimulus, may favour the localization of the stimulus, and thus facilitate the spatial effect to occur.

In Experiments 2 and 3 the detection of SC-effective (red) targets – but not SC-ineffective (blue) targets – displayed a spatial congruency effect (i.e., the MRE was larger when AV stimuli were in the same position vs. different position): a larger MRE for red (SC-

effective) than blue (SC-ineffective) targets in the multisensory *same position* condition (i.e. when the visual and auditory stimuli originated from the same spatial position); a larger MRE in the same spatial position than different spatial conditions for red targets only.

My experiments report additional evidence on the crucial role of the SC in MI. Moreover, they also contribute to clarify the role of the attention in MI, and to provide directions for future investigations on this topic. Indeed, the fact that similar pattern of results are obtained regardless of the direction of cross-modal attention is in accordance with the idea that the MI take place in an early pre-attentive stage (Vroomen et al., 2001b). A further speculations on the role of attention in MI will be provided in the General conclusion chapter.

The second contribution regards the validity of the Tritanopic technique in the study of MI and SC mediated effects.

Limitations of the method

Unfortunately, although the results may seem interesting and in accordance with my hypothesis, I have to admit a few methodological weaknesses, which became apparent to me, as a consequence of the learning process during my Ph.D. For example, SC-effective and SC-ineffective stimuli should be equated according to a fixed detection value (e.g., d'= 1), referred to the peripheral locations actually used, and for each participant.

Moreover, to validate the application of the S-cone method to every participant, one should obtain the spectral emission curves for stimuli adjusted by participants and subsequently, using cone-contrast functions, calculate the intensity of S, M, L signals.

In fact, there is a problem with using RGB channels for the creation of "pure" colour stimuli for experimental purposes. Pure wavelength stimulus, for example pure red (R= 255, G= 0, B= 0), does not stimulate only one cone type, as one would expect. An RGB picture of a natural scene has channel correlation information of the natural environment but it is in a different form than what the biological eye would see in nature. Therefore, in order to control the activation of each cone, the RGB image needs to be converted into

LMS activation values. For this reason, even if I have adopted different precautions in order to minimize the residual activation in different cones (e.g. using the yellow adapting field, or the luminance noise), the exact calculation of individual LSM activations would represent a better control of the effect of the stimuli on each singular cone type in each individual participant.

CHAPTER 3

CROSS-MODAL SHAPE CONGRUENCY EFFECT

Chapter 3:
Cross-modal shape congruency effect

Chapter overview. The aim of the present chapter is to investigate other factors that, beside proximity in space and time, can influence the strength of MI. In Chapter 2 I explored the spatial rule mediated by the SC, and different methods that can be used for the indirect investigation of the role of the SC. The influence of spatial and temporal factors on MI has been a research focus for many years (Calvert, Spence, & Stein, 2004; Holmes & Spence, 2005; Sarko et al., 2012). Recently other factors have been described. Cross-modal perception refers to all those features from different sensory modalities that can be abstracted and matched across modalities (e.g. brightness, size, loudness, pitch, lightness etc.). The facilitative effect of CMC on human performance has been already proved. However, the level of processing at which this facilitation occurs is still a matter of debate. The *speeded classification paradigm*, often used in this field, does not allow researchers to conclude whether the influence occurs at a perceptual or decisional level.

In the following chapter, two experiments will try to put light on the relation between Multisensory Integration (MI) and Cross-modal congruency (CMC). The two research fields have been usually studied separately, with different research paradigms: the speeded detection task for MI and the speeded classification task for CMC. In order to investigate the relation between these two effects, the participants in my experiments were required to perform a simple detection task with cross-modally congruent and incongruent stimuli, and the relation between the congruency effect and the inverse effectiveness rule has been investigated. Moreover, in order to avoid the involvement of *dimensional features* that can be linked by a common code for magnitude (like loudness and brightness; Marks, Ben-Artzi, & Lakatos, 2003), I selected *structural features* (i.e., shape), assuming that visual contours and sound profiles can be matched according to space-time mapping (as described in the method section of Experiment 4).

In Experiments 4-6 irrelevant sounds (spiky *tzk* vs. soft *wow*) and visual targets (spiky vs. curvy shapes) with different contrasts were presented to participants in congruent (tzk

+ spiky; wow + curvy) vs. incongruent (tzk + curvy; wow + spiky) conditions. Beside the congruency effect, the inverse effectiveness rule of MI was also investigated, under the expectation that the multisensory response enhancement decreases as the luminance contrast increases.

Similarly to experiments in Chapter 2, all the tasks in Chapter 3 were performed while attention was focused on visual stimuli and explicit responses to auditory stimuli should be inhibited. According to the results from Chapter 2, MI occurs even in a task in which the auditory stimulus is task irrelevant (and is actually part of the catch trial). Based on this result, I assumed that a congruency effect would be higly automatic, given the rapid onset, and thus occurring pre-attentively.

Chapter 3 includes an overview of the current literature on cross-modal perception, followed by two experiments. The literature overview describes early results obtained with explicit paradigms and more recent results obtained with specific paradigms (e.g. speeded classification), a summary of different types of cross-modal correspondences, as well as the description of the neuronal substrates involved in cross-modal perception. Different fields of intersection with cross-modal perception are also briefly discussed, such as sensory substitution (mapping from one sensory modality to another sensory modality, mostly with the aim of helping handicapped people). Then, two experiments are described, in which the structural congruency between visual stimuli and sounds was investigated using a simple detection task.

3.1 Introduction

Our perception of objects and events includes preferences for specific cross-modal pairings, like small visual size and low pitch, or light colour and high pitch (Ward, Huckstep, & Tsakanikos, 2006). Some cross-modal correspondences are also reflected in language, where words with a well-defined meaning in a given modality are used to describe attributes in a different modality, like when we say that the taste of a wine is sharp, or that a voice is soft. People share implicit mappings across the senses. But what are the rules that determine the cross-modal binding of perceptual features?

The interest of researchers has traditionally been focused on the role of spatial and temporal factors of MI; i.e., to extrinsic properties of multisensory stimuli. Only in recent years, it has been extended to the correspondence of intrinsic features of multisensory stimuli. Nowadays, a large body of research shows that people are sensitive to featural CMC: high pitch sounds are usually matched with small objects, a high elevation in space and a high level of brightness; vice versa, low pitch sounds are matched with big objects, a low position in space and a low level of brightness.

In the upcoming paragraphs a brief note will be dedicated to the terminology that I will adopt in the present Chapter 3, followed by an explanation of the concept of cross-modal perception, and a review of the most representative studies in the literature.

3.2 Cross-modal congruency

Two main notions will be used in the current chapter, and it is important to distinguish between the different processes to which they will refer: Multisensory integration (MI) and Cross-modal perception (or cross-modal correspondences/congruency, CMC). MI, as already defined in the main introduction, is modulated by the degree of spatial and temporal correspondence of stimuli from different modalities, which affects their interaction at the neuronal level. MI does not refer (at least in my thesis) to the influence of qualitative or semantic features of the stimuli. On the

contrary, CMC refers to the interaction between qualitative features (e.g., luminance, brightness, colour, loudness), which may happen at different stages of perception.

Both mechanisms are part of the more general cross-modal binding problem, and may be part of the same not fully understood process.

Cross-modal binding is a broad problem in cognitive neuroscience, referring to how different unisensory stimulus features are integrated into a coherent, multisensory object representation.

Many authors link CMC to synaesthesia, a peculiar condition in which, for some observers, stimulation in one modality gives rise to an automatic perceptual experience in a second modality. The synaesthetic condition may occur intra-modally (letters and numbers induce colour experience in vision; Ramachandran & Hubbard, 2001), and cross-modally (Ginsberg, 1923).

The relation between synaesthesia and the "normal" cross-modal perception is still unclear: some view synaesthesia not as an anomalous phenomenon, but rather as one reflecting a normal mode of cognition (for a discussion, see Sagiv, 2004) that remains implicit in most individuals. On the contrary, other authors claim that synaesthesia is the result of atypical connections (Hubbard & Ramachandran, 2005), or of the atypical use of normal perceptual mechanisms (Grossenbacher & Lovelace, 2001).

However, at present nothing is known about possible structural differences between the brains of synaesthetic and other members of the population, with the possibility that adult synaesthetes have retained some pathways/processes that most other members of the population lost during development (Baron-Cohen, Burt, Smith-Laittan, Harrison, & Bolton, 1996; Maurer, 1997).

As a result of the unclear relation between synaesthesia and "normal" cross-modal perception, the second one is frequently addressed in literature as synaesthetic correspondence (Parise & Spence, 2008). This term will be, however, avoided in the present chapter.

3.2.1 Early studies on cross-modal perception

The existence of cross-modal similarities among different modalities has been described early in the literature. Sapir (1929) presented visual objects of different size (small circle and big circle) and two nonsense words “mal” and “mil”, and observed that most of the subjects associated the words “mil” and “mal” with the small and big object respectively. A similar example was given by Köhler (1929), who presented two shapes (a spiky and a rounded shape) and two nonsense words (“takete” and “baluma”). Similarly to the first experiment the majority of subjects associated the word “takete” with the spiky shape and the word “baluma” (“maluma” in later editions, Spence 2011) with the rounded shape, proving that visual and auditory stimuli share some abstract proprieties that are associated in a non-arbitrary way. Ramachandran and Hubbard (2001) repeated Köhler’s experiment using the words “bouba” and “kiki”. The takete/maluma (or kiki/bouba) effect has implications for the evolution of language, since it suggests that the naming of objects is not completely arbitrary. The rounded shape may be associated to the name "bouba" because the mouth makes a more rounded shape to produce that sound while a more angular mouth shape is needed to make the sound "kiki". The sounds of a K are harder and more forceful than those of a B, as well. These and similar phenomena define a field called “phonetic (or sound) symbolism” (Brown, 1958; Imai, Kita, Nagumo, & Okada, 2008; Parise & Pavani, 2011).

Stevens and Marks (1965) demonstrated that both adults and 5-year old children match light greys with loud sounds and dark greys with quiet sounds, proving the early development of the cross-modal mapping between brightness and loudness.

Early studies, however, were primarily conducted using explicit paradigms. Only more recent studies, described in the following paragraph, utilised implicit tasks to discover CMC effects.

3.3 Cross-modal perception in the speeded classification paradigm

In recent years, behavioural studies on CMC usually adopted the speeded classification paradigm (SCP; Garner, 1974; Melara, 1989). The SCP was initially adopted for stimulus attributes (or "dimensions") within one single modality, to assess whether different dimensions can be processed independently of each other or whether judgements of one dimension are influenced by variations of the other.

During a typical task subjects are required to make speeded classifications or judgements about stimuli on the basis of variations of a single primary stimulus dimension. For example, they have to press a button as soon as they decide that a visual stimulus is bright or dark. Different sessions are conducted in which a second irrelevant dimension (e.g., size) is varied so that it is congruent (small size + high brightness) or incongruent (small size + low brightness) with the first dimension (Melara & Mounts, 1994). The typical measure of performance is *reaction time*. If the subject is unable to judge one dimension selectively, the interaction with the second dimension is thought to occur at some level of information processing.

The speeded classification task is part of the more general Garner interference paradigm (Garner & Felfoldy, 1970). In the Garner paradigm, two dimensions are paired, with attributes on the second dimension either varied orthogonally or held constant (Melara & Marks, 1990), giving rise to different tasks named the *baseline, correlated, and orthogonal* tasks. In baseline classification, the value on the dimension to be classified is varied while the value of the second (irrelevant) dimension is held constant. In correlated classification, the two dimensional values always vary together, such that the value on one dimension can be fully predicted from the value of the other, and can be of two different types: positively correlated or *congruent* (e.g. small size and high brightness) and negatively correlated or *incongruent* (e.g. small size and low brightness). Finally, in orthogonal classification, all possible values of the two dimensions are combined pseudorandomly, such that both the relevant and irrelevant dimensions vary across trials.

The interference that is usually observed in orthogonal task in comparison to the baseline task is assumed to reflect participants' inability to ignore variations of the irrelevant dimension (Garner, 1974) and is named *Garner interference* (Pomerantz, 1983).

If participants classify the first dimension more slowly when the irrelevant dimension is varied then it can be assumed that the two dimensions interact in perception (Melara, 1989). Garner (1974, 1976) has introduced a distinction between so called *integral* and *separable* dimensions, in order to capture the idea that some multidimensional stimuli might be processed holistically, in terms of their global properties (integral dimensions), while others tend to be processed analytically, in terms of their dimensional components (separable dimensions). According to Garner (1976), integral dimensions (such as brightness and saturation) give rise to poorer performance in the *orthogonal* task, but to a superior performance in the positively correlated (congruent) task. Such a superior performance is usually referred to as the *congruency effect*, and arises when performance is better with congruent than incongruent stimuli (see Clark & Brownell, 1976). For example, it has been shown that participants tend to classify congruent stimuli in the positively correlated task faster than incongruent stimuli in the negatively correlated task (Ben-Artzi & Marks, 1995; Clark & Brownell, 1976; Melara & O'Brien, 1987). Likewise, reaction times (RTs) to congruent stimuli tend to be shorter than those to incongruent stimuli in the orthogonal tasks (Bernstein & Edelstein, 1971; Marks, 1987; Stoffels & van der Molen, 1988; Stoffels, van der Molen, & Keuss, 1989).

In more recent years, the SCP has been used to explore the interactions between two dimensions from different modalities (Ben-Artzi & Marks, 1995, 1999; Melara & O'Brien, 1987).

In a recent review Spence (2011) summarised all the cross-modal correspondences that has been proved to influence performance in a speeded classification task to date: pitch/elevation (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Evans & Treisman, 2010; Melara & O'Brien, 1987; Patching & Quinlan, 2002), pitch/brightness (Marks, 1987), pitch/lightness (Marks, 1987; Martino & Marks, 1999; Melara, 1989a), pitch/angularity (Marks, 1987), pitch/size (Evans & Treisman, 2010; Gallace & Spence, 2006), pitch/spatial frequency (Evans & Treisman, 2010), pitch/direction of movement (Clark & Brownell, 1976), loudness/brightness (Marks, 1987).

One of the first studies with the SCP was by Bernstein and Edelstein (1971). The authors presented visual stimuli of different elevation and auditory stimuli of different pitch (both high/low), in congruent and incongruent conditions. Visual targets, presented on either side of the fixation cross (upper left/lower right for one group of participants,

and lower left/upper right for another group), were presented either alone or with the task irrelevant auditory stimuli (either high or low pitched, depending on the experimental condition). The task required the participants to discriminate the location of the visual target, and to ignore the auditory stimuli. The results showed an interesting effect: the participants were facilitated by the CMC (high elevation paired with high pitch and vice versa) only in the mixed design. Surprisingly, when the stimulus pairs were presented in a blocked design, the superiority of the congruent matching disappeared.

Ben-Artzi and Marks (1995) investigated further the congruency between elevation and pitch in an extended version of the experiment by Bernstein and Eldstein (1971), and were interested to describe whether and how interactions between pitch and visual position depend on the difference between the values of sound frequency, and on the difference between the vertical spatial positions. The subjects were presented with visual and auditory stimuli either in low or high spatial positions (visual judgement) or low or high in pitch (auditory judgement). Four pairs of auditory high and low sounds were presented (600-680 Hz; 460-820 Hz; 320-960 Hz; 180-1100 Hz) and four pairs of high and low visual positions (5, 7, 9 or 11 cm above and below). The differences in sound frequency were from 80 to 920 Hz, and in visual position from 10 to 22 cm. Both classification by position and classification by pitch were required from participants. A congruency effect was evident in both tasks, but pitch classification showed a stronger effect. Furthermore, the size of the pitch difference strongly affected classification by pitch and less strongly classification by position. The various differences between the positions of the visual stimulus, on the contrary, did not show any effect on the two tasks, suggesting a processing superiority of visual over auditory stimuli, already proposed by other authors (Melara & O'Brien, 1987).

Marks (1987a) in a detailed study with four experiments reported cross-modal interactions between: (a) dim/bright lights and dark/light colours accompanied by low-pitched/high-pitched tones; (b) low-pitched/high-pitched tones and dim/bright lights or by dark/light colours; (c) dim/bright lights, but not to dark/light colours, and soft/loud sounds; (d) rounded/sharp forms accompanied by low-pitched/high-pitched tones.

In a more recent study, Evans and Treisman (2010) adopted a slightly different approach to study the congruency between pitch and the visual features of vertical

location, size, spatial frequency, and contrast. Participants were divided in two groups, and each group had to perform a different task; either direct or indirect. The direct task was the usual *speeded classification* task in which participants should judge stimuli on the dimension that was predicted to be cross-modally congruent with the irrelevant stimulus. On the contrary, in the indirect task, the judgement had to be made on some dimension that was not correlated with the cross-modal dimensions. Evans and Treisman (2010) argued that in the direct task the cross-modal interaction can occur at any level between sensory registration (perceptual) and response (decisional), while in the indirect task the CMC is not linked to the response and therefore occurs at the perceptual level. The results replicated the pitch-spatial position CMC effect in both direct and indirect tasks, and provided novel evidence of an interaction between cross-modal features at an early perceptual level. Evans and Treisman (2010) provided the first experimental evidence of a cross-modal correspondence between pitch and spatial frequency and found no evidence of a correspondence between pitch and contrast.

In conclusion, let me point out that some dimensions are shown not to give rise to a congruency effect in the speeded classification task. Bernstein, Eason, and Schurman (1971) investigated the relation between pitch and hue, assuming a correspondence between low/ high frequency tones and low/high portions of the spectrum (determining hue). They did not find the expected facilitation of a high frequency tone on a stimulus from the high (short wavelength) end of the spectrum and of a low frequency tone on a stimulus from the low (long wavelength) end of the spectrum. The lack of interaction between pitch and hue is consistent with the idea that the light spectrum can be described, from a physical point of view, as a uni-dimensional continuum, whereas chromaticity (i.e., hue) is a two-dimensional variable. One might expect a congruency effect between stimuli attributes only if their perceptual correlates have the same dimensionality. However, negative results were reported by different authors also for the correspondence between loudness and lightness (Marks, 1987a), as well as between pitch and visual contrast (Evans & Treisman, 2010).

However, a critical point is that the classical speeded classification paradigm does not allow us to understand at which level of processing – perceptual or decisional – the interaction occurs (Melara, 1989; Walker & Smith, 1985; Spence, 2011). The automaticity of the interaction will be, therefore, the main aim of the present chapter.

3.4 Different types of cross-modal congruency effects

Three main types of CMC have been classified: statistically, structurally and semantically mediated. They may happen at different stages of human information processing and have different behavioural consequences. In particular, statistical and structural congruency might be mediated at either the perceptual or decisional level, or both, while semantic congruency might be mediated at the decisional level only.

3.4.1 Statistical

The statistical CMC is thought to be the consequence of pairs of stimulus dimensions that happen to be frequently correlated in nature. This kind of cross-modal correspondence is based on environmental regularities and may be learned. Gaver (1993) highlighted the importance of studying sounds in relation to its source, and differentiated between the listening to sounds (e.g., music, in terms of sensory qualities) and listening to events in the world (in terms of its source). In his definition of sound-event relations Gaver (1993) refers to the notion of direct perception introduced by Gibson (1979) suggesting that sounds may contain invariant properties that indicate the physical characteristics of environmental objects and their interaction, and specify a range of events that may have produced the sound. With the term *nomic mapping* the author refers to the association between sounds and non-acoustic properties of events that can physically produce them. For example, larger objects vibrate with larger oscillations and consequently produce lower frequencies than smaller objects. As such, frequency becomes an invariant indicator of size. On the contrary, the term *symbolic mapping* relies on social conventions for meaning, such that the acoustic properties of the sound are arbitrarily connected to the event it signals. For example, the sound of a fire alarm can be associated to fire flames, without being similar to the sound of burning materials.

Experimental evidence of the distinction between nomic and symbolic mapping have been provided by Coward and Stevens (2004). The authors used a speeded classification

task to study the validity of the nomic mapping between object size and frequency (pitch), opposite to the symbolic mapping between object size and the sound transformation specified by the value of the damping constant. According to Gaver (1993), frequency and size will interact and therefore, when presented together in a congruent pairing (e.g. low frequency - large object), a congruency effect will occur (better performance in congruent than incongruent conditions). On the contrary, damping usually indicates the material of an object, and has no natural relation to the size, and the relation with the size is therefore only symbolic. In such a case, it is not expected to produce a congruency effect, since the two dimensions are linked only at a post-perceptual level.

Six different types of stimuli were presented to participants: a low or high frequency sound, the drawing of a long or short bar and high or low damping constants. The stimuli were presented in congruent pairing (long bar with low frequency or low damping constant and short bar with high frequency or high damping constant) and incongruent pairings (long bar with high frequency or high damping constant and short bar with low frequency or low damping constant). Participants were instructed to classify each bar as long or short while ignoring the sound or damping constants. A better performance were present with congruent over incongruent pairings of the bar with sounds. However, no interaction between size and damping were present, proving the initial hypothesis of the authors.

3.4.2 Structural

A common belief among some authors is that cross-modal correspondences are remnants of neural connections present at birth, which influence the development of perception and language. For example, Mondloch and Maurer (2004) explored the correspondences between high-pitched sounds and small, bright visual objects in pre-schoolers (30–36 months), reporting that the majority of children matched the higher pitched sound to the lighter ball. Given the young age of the subjects, the authors concluded that CMC cannot be attributed to learning.

An alternative explanation is that CMC may occur between magnitude-related dimensions such as loudness and brightness. According to Walsh (2003) time, space and

quantity are part of a generalized magnitude system. The author proposed *A Theory Of Magnitude (ATOM)* as a conceptually new framework within which to reinterpret the cortical processing of these elements of the environment.

3.4.3 Semantically mediated

The last type of CMC occurs when people use the same term to describe stimuli in different modalities. The most common example are the words “low” and “high” used to describe both the elevation of visual stimuli and the pitch of sounds (Gallace & Spence, 2006; Martino & Marks, 1999; Melara, 1989). However, many other words can be used cross-modally: adjectives sharp, bright and dull can be applied to sounds, tastes and smells; weak and strong to almost all sensory modalities (Marks, 1982a, 1982b).

In his recent review Spence (2011) summarised the three types of CMC in regards to their causes and consequences: statistical CMC (e.g., pitch-size) probably depends on learned associations of frequent similarities from the environment, with perceptual and decisional consequences; structural CMC (e.g., loudness-brightness) is probably innate or caused by maturation of neural structure, and its influence can be at both perceptual and decisional levels; semantic CMC is learned through the development of language and its influence is mainly decisional

3.5 Neuronal correlates of cross-modal congruency

Studies that investigated the neural correlates of CMC used linguistic (Calvert et al., 1999; Raij, Uutela, & Hari, 2000; Callan, Callan, Kroos, & Vatikiotis-Bateson, 2001; Saito et al., 2005) and non-linguistic stimuli (Bushara, Grafman, & Hallett, 2001; Foxe et al., 2000; Giard & Peronnet, 1999; Raij et al., 2000).

3.5.1 Studies using linguistic stimuli

The study of cross-modal perception of linguistic events has been mainly focused on audio-visual speech production. It is well known, for example, that speech comprehension is improved when the listener can both see and hear the speaker (Reisberg & Leak, 1987) but impaired if the movements of the lips and the spoken word are semantically incongruent (Dodd, 1977). Calvert, Hansen, Iversen and Brammer (2001) tried to define the neuronal basis of this association presenting to participants congruent and incongruent lips and speech configurations. Following the three criteria established for the identification of brain areas (co-responsiveness, positive interaction in the case of congruent AV speech and negative interaction in the case of incongruent AV speech) they identified the posterior ventral bank of the Superior temporal sulcus (STS) as the substrate of cross-modal speech perception. Raij et al. (2000), using similar criteria adopted by Calvert et al. (2001) and presenting letters and meaningless stimuli to their subjects, reported an activation corresponding to the AV integration in different brain areas: left fronto-parietal region, right frontal cortex, right temporo-occipito-parietal junction and right and left STS. Among those areas, only the right temporo-occipito-parietal junction and STS area exhibited stronger activation for letters than for meaningless stimuli. However, the congruency effect (superiority of activation for congruent over incongruent sound-letter pairing) was observed only in the STS area.

3.5.2 Studies using non-linguistic stimuli

The neuronal sites that are activated during the integration of non-linguistic features have been investigated in studies that provided a detailed differentiation between cortical regions activated during congruent or incongruent AV perception. For example, Belardinelli et al. (2004) reported the activation of the left medial and temporal cortex and of the lingual gyrus of both cortical hemispheres in response to congruent vs. incongruent stimuli.

von Stein, Rappelsberger, Sarnthein and Petsche (1999) conducted an EEG study presenting different semantic stimuli to participants: object from daily life (hammer, dog etc.), written words and spoken words, in congruent or incongruent pairings. Subjects were required to perceive the meaning of the word and to let associations and imagery arise, without being requested to perform any behavioural response. The authors found that a coherent activity in the beta band between temporal and parietal lobe may play a role in the integration of semantic features originating from the same object.

Other neuropsychological data suggest that damage to the angular gyrus (in the temporo-parieto-occipital region) can interfere with the ability to match stimuli cross-modally, as assessed by the bouba-kiki test (Ramachandran & Hubbard, 2003; pairing of the words “bouba and “kiki” with spiky and rounded shapes).

Studies in primates suggest that the process of cross-modal binding takes place in the perirhinal cortex, in opposition to the STS region mostly cited in human research. However, Taylor, Moss, Stamatakis and Tyler (2006) reported that both areas play a crucial role in human information processing, with the STS acting as a pre-semantic hetero-modal region for cross-modal features, whereas only the perirhinal cortex binds those features into a high level object representation taking into account the semantic meaning of the unisensory components.

3.5 Multisensory shape perception and sensory substitution

Beside the classical studies of CMC, the field of research known as *sensory substitution* has been focused mainly on stimulus "shape". The sensory substitution technique transforms stimuli of one sensory modality into stimuli of another sensory modality, so that information gained from one sensory modality can reach brain structures physiologically related to other sensory modalities. For example, in the touch-to-visual sensory substitution, the information from touch receptors is relayed to the visual cortex, so that the person can perceive to see.

This technique, ideated mainly to help people with handicap, has provided an interesting paradigm for the study of other forms of cross-modal interactions. The main conclusion evident from these studies is the importance of *shape* as a cross-modal and transferable property, which has been studied for decades mainly in the frame of visual perception.

The neuronal site that has been pointed out as a possible neuronal substrate is the *lateral occipital cortex (LO)*. Evidence for selective shape perception activation is so robust, that some authors have defined this region as the *shape operator* of the brain (James, Stevenson, Kimb, VanDerKlok, & James, 2011). The activation of this area is shape selective with visual and haptic sensory inputs (James, Culham, Humphrey, Milner, & Goodale, 2003; James et al., 2002; James, Kim, & Fisher, 2007; Stilla & Sathian, 2008; Tal & Amedi, 2009). Area LO is also object selective with auditory inputs (Amedi et al., 2007; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; James, Stevenson, Kim, Vanderklok, & James, 2011; Lewis et al., 2004). While the evidence of a convergence of both visual and haptic information to this region is consolidated, it is not clear whether the auditory information might also contribute to this scenario. However, a recent study by James et al. (2011) reported that sound can also carry acoustic information about the shape of an object (e.g. when a sound is produced by an object's impact with a surface). The objects presented to the participants were of two shapes, rods and balls, of two different materials (metal and wood) and the auditory stimuli were created from recordings of objects impacting a hard surface. The task required from participants was to categorize the impact sounds in one of three tasks: (a) by the shape of the object; (b) by the material; (c) by using all the

information available. Area LO was more strongly activated when subjects discriminated impact sounds based on the shape of the object suggesting that this may be shape selective regardless of sensory input modality.

In current studies shape is interpreted as a complex property and an essential aspect of object recognition that can be conveyed by parallel channels from different sensory modalities. The relation between this particular field of research and the field of CMC studied with the classical speeded classification paradigm is still an open question.

3.6 Cross-modal perception and multisensory integration

Only few studies in the literature tried to solve to problem of the relation between CMC and MI. The first study was by Miller (1991), who demonstrated a congruency effect with a speeded response in a go-no go task. The participants had to respond to any stimulus, and the authors reported that participants responded more rapidly to congruent (e.g. high elevation of visual target and high pitch sound) than incongruent (high elevation and low pitch) pairs of stimuli. A more detailed explanation of the experiment will be reported in the next chapter.

The same question was addressed by Parise and Spence (2009) in a temporal order judgement task: given two stimuli, one auditory and the other visual, participants should provide an un-speeded judgement regarding their order (auditory first or visual first). The authors investigated the relation between pitch and size in the first experiment, and the relation between pitch and shape in the second one. The stimulus onset asynchrony (SOA) was varied on a trial-by-trial basis with the method of constant stimuli, and the pairings of stimuli were either congruent or incongruent. They assumed that, if CMC operates at perceptual level, it should be harder for participants to treat the two stimuli as being separate and thus to detect small temporal differences among them.

The results confirmed such a hypothesis: the *just noticeable difference* (JND) was significantly higher for congruent that incongruent stimuli. In a previous study Parise and Spence (2008), the same hypothesis was studied in the *temporal ventriloquism illusion* (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). The term temporal ventriloquism refers

to an illusion in which audition affects vision on the temporal dimension, and it has been shown to be stronger in the case of congruent cross-modal stimuli. Parise and Spence (2008) further demonstrated that the same influence is evident also in the case of *spatial ventriloquism*; that is, congruent pairs of stimuli give rise to a significantly larger illusion than do incongruent pairs.

3.8 Cross-modal shape congruency in a focused attention paradigm

3.8.1 Rationale of the experiments

The MI, resulting in a multisensory response enhancement (MRE), has been studied in the present chapter in relation to the congruencies between features coming from different modalities (visual and auditory).

The previous chapter has described a few studies that tried to clarify the relation between MI and CMC. However, with the only exception of Miller (1991) the conclusions were always based on indirect assumptions (e.g., based on temporal order judgement task). Here I argue that, in order to prove that cross-modal perception and multisensory integration are part of the same mechanism, MI and CMC have to be studied within the same paradigm. MI has usually been studied with speeded detection responses to simple stimuli, and the usual manipulations were related to the spatial and temporal coincidence between visual and auditory stimuli, or the saliency of the stimuli. On the other hand, CMC has been usually studied with the speeded classification paradigm, manipulating the congruency between different features. In order to explore the relationship between these two effects, congruency has been investigated within the paradigm of the MI, with a simple detection task (instead of the speeded classification, usually adopted).

The two main predictions of the experiments of the present Chapter 3 were, thus, the following:

- I expected the responses to be faster in the multisensory condition (when a sound and a visual shape are presented together) than in the unisensory condition (when only a visual shape is presented, either spiky or curvy); and the advantage should be larger in congruent (spiky sound coupled with a spiky visual shape, curvy sound coupled with a soft visual shape) than incongruent (spiky sound coupled with a soft visual shape, curvy sound coupled with a spiky visual shape) conditions.

- If the two processes are part of the same mechanism of MI, the amount of cross-modal shape congruency should follow the inverse effectiveness rule (Experiment 4). This will be tested introducing visual stimuli of 4 different luminance contrast (grey colour) with the background (black colour). In this way, higly efficient stimuli (high luminance contrast with the background) are expected to exhibit the weaker MRE, and the amount of MRE is expected to be inversely linear to the luminance contrast of stimuli.

In comparison to other studies on CMC, the present study is important for a few reasons:

- First of all, it is the first experiment that investigates the CMC of a complex attribute like visual and auditory shape, in a speeded detection task. In fact, even if the shape of visual stimuli has been studied in previous studies (Miller, 1991), the auditory shape has never been manipulated in a speeded detection task so far, perhaps because of a less intuitive definition of what is here called *auditory shape*. Similar stimuli pairings were, however, used by Parise and Spence (2009, Exp. 2) in temporal order judgement task. In my experiments, visual shape (spiky vs. curvy) consists in changes of contour directions in space; auditory shape (tzk vs. wow sounds) consists in changes of intensity and number of interruptions in time (further explanation below).
- A congruency effect in speeded detection would indicate that the association is based on the structural similarity of AV stimuli, probably operating at a low level of information processing.

In conclusion, the logic underlying the creation of the stimuli is as follows. The common feature between auditory and visual stimuli was, as already mentioned, the *shape*. The term may have a broad and complex meaning, and a better definition is therefore needed. When we talk about visual stimuli, a definition is more obvious, and relates to some properties of the *contour*. In the case of the stimuli that I used, a *spiky shape* is defined by frequent (12 cycles) and sharp discontinuities of the contour, while the *curvy shape* is defined by infrequent (3 cycles) and smooth changes of direction of the contour (Figure 3.1 a). In the case of the auditory stimulus, the definition is less intuitive. The spiky *tzk* sound is characterized by frequent and abrupt interruptions of intensity of the sound, while the soft *wow* sound is characterized by infrequent and smooth changes of

intensity, without clear interruptions (Figure 3.1 b). Similar terms have been studied and described in the literature. For example, in the study by Weinberger & McKenna (1984) *tonal contour* was defined as an acoustic sequence that increases or decreases monotonically in frequencies, or a non-monotonic sequence. The same term is linked to the study of some particular tonal languages (e.g. Chinese) in which there are shifts from one pitch to another over the course of the syllable or word. More broadly speaking, the term *acoustic contour* was used by Schmuckler and Gilden (1993) to define noises varying in pitch, duration and intensity of the notes in a sequence. Another similar term is the *auditory temporal edge* in the study by Chait, Poeppel, & Simon, (2008), conceptually defined as transitions between constant tones, regularly alternating, and randomly alternating tone-pip sequences.

In my experiments, the term *shape* is primarily defined by the interruptions in the sequence of frequencies varying in intensity, which constitutes the main manipulation adopted in order to differentiate the *spiky* from the *soft* sound. As such, the term *auditory shape* agrees only partially with the other terms above described. Even if the spiky and soft sound are constituted by variations in frequencies (see waveforms and spectra of the stimuli in Figure 3.1), they do not represent the main manipulation. In fact, the term that more closely resembles the auditory shapes used in my experiments, are the so called *amplitude transients*, studied by Fishbach, Nelken and Yeshurun (2001), in an attempt of linking the auditory edge detection to the visual edge detection. This study will be further discussed in Chapter 6.

It is important to note, however, that the manipulation in the case of the visual stimulus (spiky/curvy) is related to the spatial domain, while in the case of the auditory stimulus the changes are related to the temporal domain. Further explanation of this point will be given at the end of the chapter, in the general conclusion section.

3.8.2 Experiment 4

3.8.2.1 Method

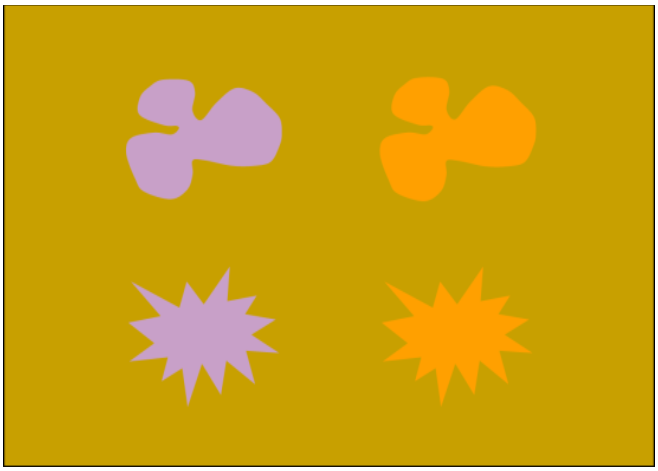
Participants

Eleven healthy right-handed students (age range 20 - 26 years, 4 males, 7 females) took part in the experiment. All had normal hearing and normal or corrected-to-normal visual acuity. They were naïve as to the purpose of the investigation, gave their informed consent prior to the beginning of the experiment, and received course credit for their participation.

Apparatus and stimuli

Stimuli were generated by a computer equipped with standard graphic and acoustic software. Superlab 4.8 was utilized to present stimuli and collect responses. Visual stimuli were displayed on a Sony Trinitron CPD-G200P CRT 17 inch monitor, set at 50% brightness, 90% contrast, 1280 $\times$ 768 pixel resolution, 60 Hz. The participant was comfortably seated in a dimly illuminated cubicle at a distance of 57 cm from the centre of the screen and required to fixate a central white cross (70 cd/m²) against a black background (5 cd/m²). Visual shapes used as targets are shown in Figure 3.1. They were regions bounded by spiky or curvy contours, subtending 3.0° horizontally and 2.7° vertically, lighter than the background.

A.



B.

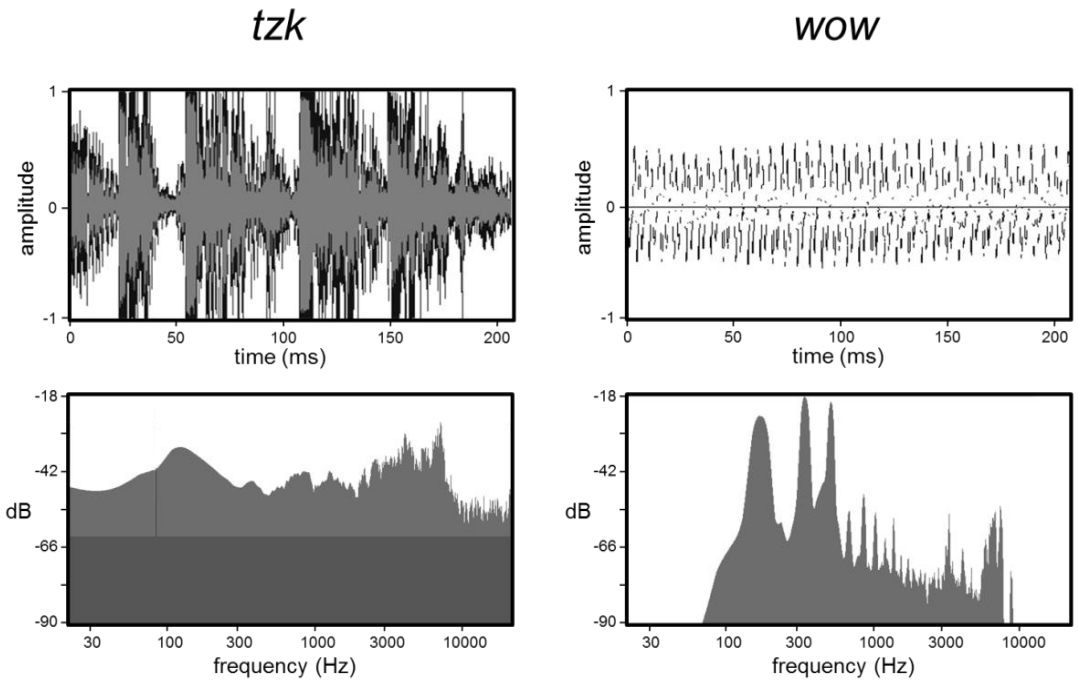


Figure 3.1 Experiment 4, visual and auditory stimuli. A) Spiky and Curvy shape (adapted from Köhler,1929), in blue "low saliency" (on the left) and red "high saliency" (on the right) condition, on a yellow adapting background (see Experiment 3, Chapter 2 for detail on the methodology). B) Waveforms (upper graphs) and spectra (lower graphs) of the tzk and wow sounds used in Experiments 4, 5 and 6.

Visual stimuli were either red or blue spiky and curvy shapes (Figure 3.1 A.) on a yellow background. The values for red are the same as in Experiment 3 from Chapter 2

(based on the tritanopic technique): a blue stimulus on a yellow adapting background should carry only colour contrast information and be “lowly efficient”, while the red stimulus should produce a strong input to the luminance pathway (and thus carry luminance contrast information) and be “highly efficient”. The saliency is defined in relation to both contour information and RT. In fact, RT to S-cone isolating stimuli are notably longer (e.g. O'Donell & Colombo, 2006), and therefore I expected faster responses to red than blue visual shapes. Moreover, since the information about edges is carried mostly by the luminance pathway (red stimuli), I assumed that a shape congruency effect will arise with red stimuli, and will not arise with blue stimuli.

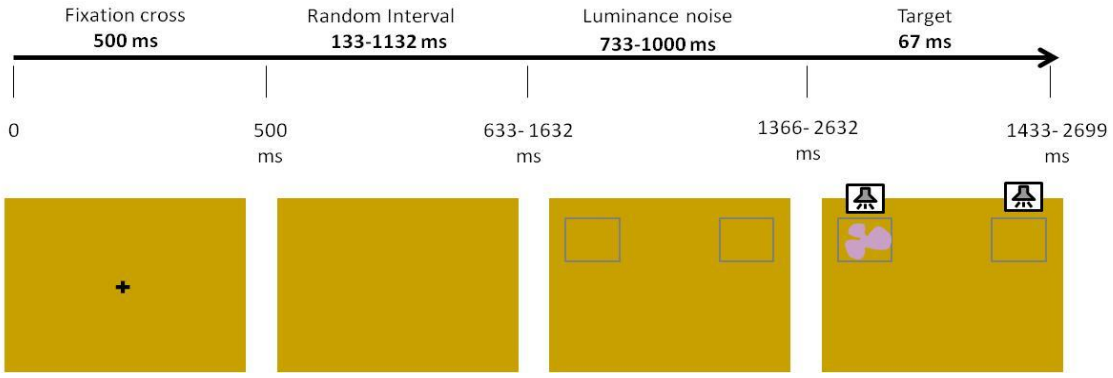


Figure 3.2 Experiment 4, stimuli and timetable of a typical trial. Each trial was initiated by a fixation cross lasting for 500 ms, followed by a variable interval (lasting from 133-1132 ms) and two outline squares appearing on the right and left corner at the top of the screen, with a variable duration from 733 – 100 ms.The target was the spiky or curvy shape, in the red (high luminance contrast) or blue (low luminance contrast) condition, unisensory or multisensory (AV) condition. Loudspeakers were placed just above the outline squares. In multisensory positive trials an auditory stimulus was presented simultaneously with the visual target, in congruent or incongruent combination.

The acoustic stimuli were 210-ms sounds (*tzk* and *wow*) selected to match the two visual shapes. The *tzk* sound was a discontinuous sequence of metallic bursts; the *wow* sound resembled a continuously modulated trumpet tone. Their waveforms and spectra are illustrated in Figure 3.1 b. In training and experimental trials sounds were presented by loudspeakers located above the upper corners of the computer screen, where visual targets were presented. The position of the acoustic source (left vs. right loudspeaker) was

always consistent with the position of the visual target, favouring MI on the basis of the spatial rule.

Procedure

Prior to Experiment 4, a preliminary study was conducted on a small group of different participants (N= 8). They were all asked to verbally associate the two sounds (*tzk* and *wow*) with the two visual stimuli (spiky and curvy shapes shown in figure 3.1). All participants paired the *tzk* sound with the spiky shape and the *wow* sound with the curvy shape. This preliminary study indicated that the explicit associations between the two pairs of congruent stimuli were strong.

The participant was instructed to press the left button of the mouse with the index finger of the dominant hand as quickly as possible to respond to the presentation of a visual target (either spiky or curvy, in either unisensory or multisensory conditions), while suppressing any overt response to the sound alone. Since attention should be concentrated on visual stimuli, acoustic stimuli were task-irrelevant.

All trials included three events: 500-ms background with central fixation cross; empty background lasting 133-1132 ms; two outline squares, lasting 733-1100 ms, at the left and right of the fixation cross, marking the possible locations of the visual target. Stimulation conditions differed according to the last 67-ms event, which included a spiky/soft sound in Unisensory auditory (catch) trials; a red/blue spiky/curvy shape in Unisensory visual trials; a spiky sound coupled with a spiky shape or a soft sound coupled with a curvy shape in Multisensory Congruent (MSC) trials; a soft sound coupled with a spiky shape or a spiky sound coupled with a curvy shape in Multisensory Incongruent (MSI) trials (Figure 3.2).

Each participant was shown 480 trials, divided in three sets of positive trials (unisensory visual USV, multisensory congruent MSC, multisensory incongruent MSI) and a set of catch trials, divided in 15 blocks. Every block contained 32 trials in a different random sequence, including: 8 catch trials (resulting from the 2×2×2 combination of position, sound and repetition), and 8 US, 8 MSC, and 8 MSI trials (all resulting from the 2 × 2 × 2 combination of colour, position and shape).

After verbal instructions and 20 min of dark adaptation the participant entered a 38-trial training session in which three criteria should be reached: a maximum of two RTs over 700 ms, a maximum of two RTs under 120 ms, and a maximum of two errors (either misses or false alarms). If the participant failed to meet any of the three criteria, the training session was repeated. The participant would be eliminated if he/she failed again during the supplementary training session (which did not occur). After a brief rest the participant entered the experimental session containing the above described 360 trials, presented in a random sequence different for every participant.

3.8.2.2 Results

In first place, the analysis was performed on the level of accuracy of the participants, expressed as proportion of misses and false alarms. The proportion of false alarms was 1%. Similarly to the analysis in the previous chapter, a preliminary test was conducted of the distribution of Misses against zero, on transformed data [x= arcsin $(pMiss)^{0.5}$]. The analysis showed a significant difference in only 3 cases, all in the US condition (t values larger than 4, df= 10, one-tailed, p smaller than the 0.004 critical value chosen after the Bonferroni correction). Despite only a few Misses values were different from 0, they were all in the US condition. The ANOVA reported a significant effect of the factor Condition $F_{2,20}$= 21.3 p< 0.0001, but no effect of the factor Shape $F_{1,10}$= 1.86 p= 0.2, nor of the factor Saliency $F_{1,10}$= 0.94, p= 0.35, and no effect of the interactions: Condition x Shape $F_{2,20}$= 0.316, p= 0.73, Condition x Saliency $F_{2,20}$< 1, Shape x Saliency $F_{1,10}$< 1, and Condition x Shape x Saliency $F_{2,20}$< 1. This analysis indicated that overall there were a reduction of the level of Misses in both the MSC and MSI conditions in relation to the US condition (US vs. MSC= 4.2% vs. 1.6%, t(10)= 5.2, p< 0.0001; US vs. MSI= 4.2% vs. 1.2%, t(10)= 9.4, p< 0.0001).

First, I performed a planned comparisons analysis to evaluate the expected effects of multisensoriality, congruency and saliency. As regards multisensoriality, participants were faster in MSC and MSI conditions than in the US condition (US= 331 ms, MSC= 309 ms, MSI= 315 ms; US vs. MSC, t(10)= 4.3, p< 0.001; US vs. MSI, t(10)= 3.5, p< 0.005). As

regards the congruency effect, the difference between MSC and MSI means (6 ms) was in the expected direction, although it did not reach the significance level ($t(10)= 1.9$, one tailed, $p= 0.09$). As regard saliency, participants were faster with $High_{red}$ than Low_{blue} targets (310 vs. 326, $t(10)= 6.8$, one-tailed, $p< 0.001$).

Then, I performed a within-subjects ANOVA on the whole set of data, with Condition (US, MSC, MSI), Saliency (Low_{blue}, $High_{red}$) and visual Shape (spiky, curvy) as factors. All main effects and interactions were significant. Confirming the result of the planned comparisons analysis, the main effect of Condition ($F_{2,20}= 10.5$, $p< 0.001$) and Saliency ($F_{1,10}= 38.1$, $p< 0.001$) were significant.

Unexpectedly, there was also a difference between the two visual shapes, with participants being overall faster in responding to the curvy that spiky shape (309 vs. 327 ms, $F_{1,10}= 50.8$, $p< 0.001$).

In conclusion, all the interactions were also significant: Condition x Shape $F_{2,20}= 6.71$ $p< 0.01$, Condition x Saliency $F_{2,20}= 15.0$ $p< 0.0001$, Shape x Saliency $F_{1,10}= 28.4$ $p< 0.0005$, Condition x Shape x Saliency $F_{2,20}= 6.94$ $p< 0.005$.

Given the main effect of shape, the interpretation of all interactions will be performed on the basis of two ANOVAS on spiky ad curvy shape separately, as follows.

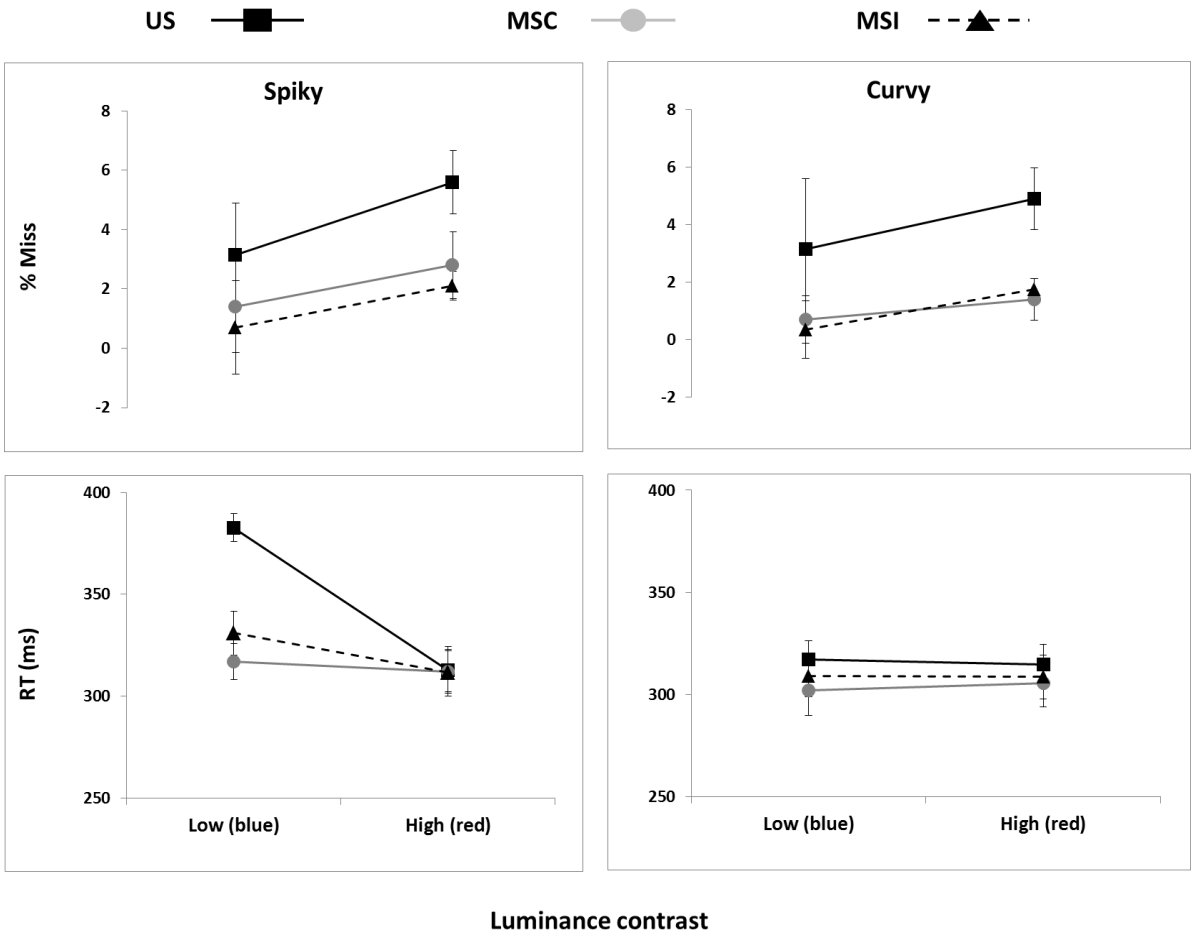


Figure 3.3 Experiment 4, % Miss and RT. % Miss (upper graphs) and RTs (lower graphs; ± 1 s.e.m.) for the Target (2) × Condition (Lee & Stromeyer, 1989) × Contrast (2) design of Experiment 4. Detection was clearly harder when both targets were in the US condition. In general, performance in multisensory congruent (MSC) and multisensory incongruent (MSI) trials was better than in unisensory trials, both for the accuracy and RTs, with the only exception of the curvy shape, where no difference was evident for the RTs.

The main analysis on the congruency effect was performed on the %MRE measures (following the formula adapted for the focused attention paradigm in Chapter 2), separately for the two shapes. Figure 3.4 report the %MRE for curvy and spiky shape, for the Low_{blue} saliency and $High_{red}$ saliency stimuli.

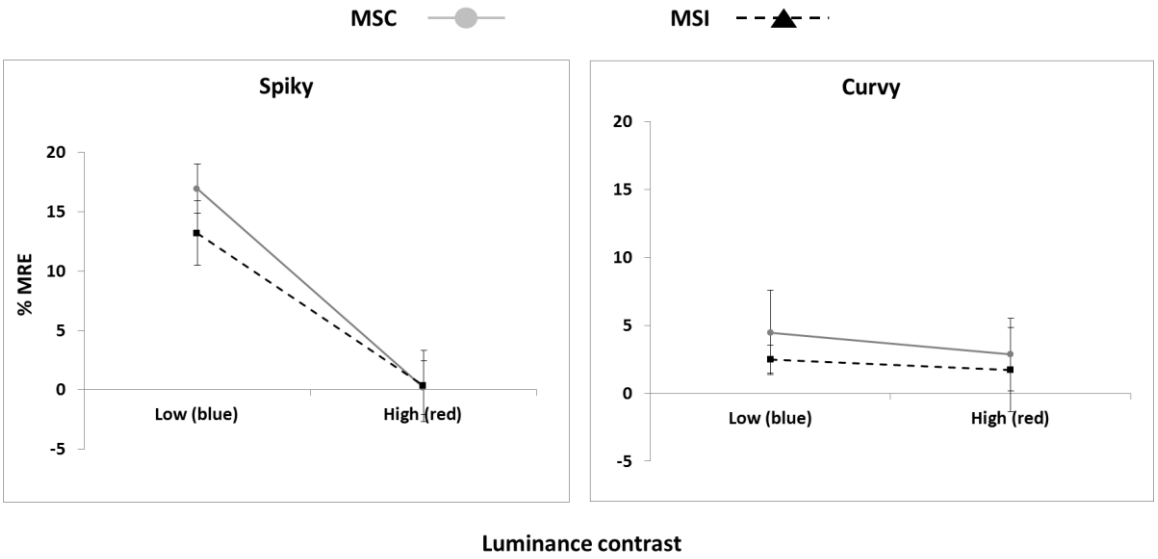


Figure 3.4 Experiment 4, % MRE (± 1 s.e.m.)**.** The % MRE was calculated for the spiky and curvy shape separately, in the MSC and MSI condition, for the low luminance contrast (blue) and high luminance contrast (red) stimuli. Overall the %MRE was stronger in the spiky shape, and for the blue stimuli. For the spiky shape a clear superiority of both MSC combinations over MSI was present. This pattern of results, however, was not present in the case of the curvy shape.

A series of t test against 0 revealed that only two conditions out of eight presented a %MRE significantly different from 0: $MSC_{blue\text{-}spiky}$, $t(10)= 9.27$, $p< 0.001$; $MSI_{blue\text{-}spiky}$, $t(10)= 5.55$, $p< 0.001$; whereas the %MRE for all other combinations did not differ from 0: for $MSC_{red\text{-}spiky}$, and $MSI_{red\text{-}spiky}$, $t(10)< 1$; $MSC_{red\text{-}curvy}$, $t(10)= 1.59$, $p= 0.1$; $MSI_{red\text{-}curvy}$, $t(10)= 1.23$, $p= 0.2$; $t(10)= 1.94$, $p= 0.08$; $MSC_{blue\text{-}curvy}$, $t(10)= 1.94$, $p= 0.08$; $MSI_{blue\text{-}curvy}$, $t(10)= 2.9$, $p= 0.02$; for the critical p value of 0.01 chosen after the Bonferroni correction for multiple comparisons).

As expected on the basis of the RT analysis, the pattern of results was different in spiky and curvy conditions. Two different Saliency (low, high) x Congruency (MSC, MSI) within-subjects ANOVAs were run for spiky and curvy shape separately. The result for the spiky shape indicated a main effect of Saliency ($F_{1,10}= 29.6$, $p< 0.001$), while the main effect of Congruency ($F_{1,10}= 2.28$, $p= 0.16$) and the Saliency x Congruency interaction ($F_{1,10}= 1.53$, $p= 0.24$) were in the expected direction but did not reach the statistical significance level. The effect of Saliency was clearly due to the superiority of the %MRE in Low_{blue} over $High_{red}$ saliency condition, in both MSC (MSC_{red} vs. MSC_{blue}, $t(10)= 6.13$ $p< 0.001$) and MSI (MSI_{red}

vs. MSI_{blue}, t(10)= 3.67, $p< 0.01$) conditions, indicating that the improvement is stronger for weak stimuli, according to the inverse effectiveness rule of MI.

A planned comparison analysis, motivated by the existence of strong predictions, revealed a significant difference between MSC and MSI conditions for blue target in the spiky condition (MSC_{spiky_blue}= 16.9% vs. MSI_{spiky_blue}= 13.2%, t(10)= 2.9; $p< 0.05$). As regards to curvy shape, no significant effect was found for the factors Saliency $F_{1,10}< 1$, Condition $F_{1,10}< 1$, or Saliency x Condition interaction $F_{1,10}< 1$.

The absence of any effect with the curvy shape may be explained by an outstanding floor effect present in all four conditions. Since the stimulus was optimally efficient in the US condition, the multisensoriality did not confer any additional enhancement. The floor effect with $high_{red}$ saliency stimuli is understandable, given the high luminance contrast evoked. It is less intuitive, however, the floor effect with the blue stimuli, especially if we take into account the absence of such a pattern with the spiky shape. We could speculate that the curvy shape were more efficient than the spiky shape, supposedly for some reasons related to the contour. If we consider the spiky shape as being fuzzy, it may have resulted in a less prominent luminance contrast elicited by its fuzzy contour, while the continuity of the curvy shape resulted in a strong luminance contrast elicited by the contour information, giving rise to the difference in the efficiency even in the low_{blue} saliency condition.

3.8.2.3 Conclusions

The results obtained in Experiment 4 are weak and do not support any strong conclusion. Overall, it seems that the saliency of the stimuli was too high, and the performance of the participants was already at an optimal level, resulting in a *floor effect*, demonstrated by the fact that a significant %MRE was elicited only in two out of eight conditions (MSC_{spiky_blue} and MSI_{spiky_blu}). Therefore, the only condition that seemed to benefit from the multisensoriality overall, and from the congruency in addition, is the spiky shape paired with the spiky sound in the low_{blue} saliency stimulus condition. A significant %MRE in this condition is expected on the basis of the IER, which states that the weaker is the response in the US condition, the stronger is the MRE in the MS condition. However,

the existence of a shape congruency effect (although very weak) is in contradiction with the initial hypothesis of an absence of a shape perception with low_{blue} (equiluminant) stimuli. This unexpected result can be explained in two ways: either I have to suppose that I did not manage to reproduce a perfect equiluminant condition, and therefore the blue "low saliency" stimulus elicited a response in the luminance channel (mediating shape perception). Alternatively, I should speculate about shape perception mediated by the colour contrast elicited by the low_{blue} stimulus, as already proposed by other studies (Gregory, 1977; see Cavanagh, 1991 for a review). In fact, the assumption that colour contrast does not mediate the perception of shape is incorrect and superficial, and many authors in the past have claimed the opposite (Favreau & Cavanagh, 1981; Flanagan, Cavanagh, & Favreau, 1990; Bradley, Switkes, & De Valois, 1988; Cavanagh, Arguin, & Treisman, 1990). Any further speculation on the argument may be superfluous given the weakness of the results.

Overall, I have to admit the inadequacy either of the experimental design or of the stimuli used in Experiment 4 in eliciting the main shape congruency effect. However, I took the data present here as an indication of a possible effect, deepened in the following Experiment 5.

A general methodological conclusion is that, in order to investigate multisensory effects, and especially congruency effects in simple detection task, one should try to avoid creating highly efficient stimuli in unisensory condition. In fact, if the response of the subject is already optimal in the US condition, the MI is not needed anymore. Therefore, in order to encourage the MI effects to occur and the congruency effects in addition, we should try to create non optimal conditions in US trials, either with lower luminance contrast, or abbreviating the duration of the stimuli.

3.8.3 Experiment 5

In Experiment 4 I failed to elicit any strong effect. However an indication of a possible shape congruency effect motivated me to run a second experiment. The pattern of data seemed to depend on the saliency of the stimuli. However, two levels of saliency (i.e. $high_{red}$ and low_{blue}) were not sufficient to clarify the pattern of improvement related to

the saliency, especially if we take into account the fact that we may have been in front of a *floor effect* in the case of the $high_{red}$ saliency target.

In addition, I was interested in describing the function that governs the *inverse effectiveness rule.* In Experiment 5 I introduced a few modifications of the previous stimuli and design. In fact, stimuli were grey on a black background, with four levels of luminance contrast (Figure 3.5). By introducing grey achromatic stimuli instead of colour stimuli it was easier to control their respective effectiveness. In fact, while the influence of the luminance contrast or colour contrast on RT are well described by the Pieron's law (see also Chapter 2, p. 54), their combined effects are still unclear (Hsu, 2005). In order to have a better control of our predictions, the efficiency was defined as the luminance contrast of the stimuli with the background, and the colour was omitted in the present experiment.

3.8.3.1. Method

Participants

Eighteen healthy right-handed students (age range 19-27 years, 5 males and 13 females) took part in the experiment. All had normal hearing and normal or corrected-to-normal visual acuity. They were naïve to the purpose of the investigation, gave their informed consent prior to the beginning of the experiment, and received course credit for their participation.

Apparatus and stimuli

The same visual shapes from Experiment 4 were used in Experiment 5. In order to eliminate every possible influence of colour, in Experiment 5 only grey stimuli were presented, with different luminance contrasts with the background (figure 3.4). Four luminance levels were used to modulate the salience of visual targets (5.10, 5.56, 7.61, 22.46 cd/m²) corresponding to the following Michelson contrast values: 1.0, 5.3, 20.7,

63.6%. The same auditory stimuli from Experiment 4 were used in Experiment 5 (Figure 3.5).

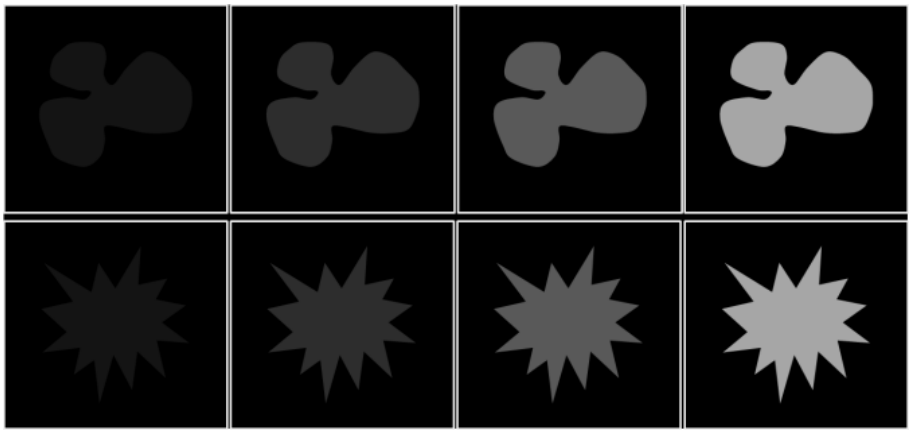


Figure 3.5 Experiment 5, visual stimuli. Spiky and curvy shapes used as targets in Experiment 5 (adapted from Köhler, 1929), in 4 different luminance contrasts.

Procedure

The experimental session included three sets (unisensory visual USV, multisensory congruent MSC, multisensory incongruent MSI) of positive trials, each including two subsets of 48 trials (6 for every target combination of 4 luminance contrasts and left/right spatial positions) and one subset of 72 catch trials, as follows.

USV trials: 48 with the spiky shape and 48 with the curvy shape;

MSC trials: 48 in which the spiky shape was paired with the tzk sound and 48 in which the curvy shape was paired with the wow sound;

MSI trials: 48 in which the curvy shape was paired with the tzk sound and 48 in which the spiky shape was paired with the wow sound;

Catch trials: 18 (unisensory auditory) for each combination of sound (tzk/wow) and position (left/right), in the absence of visual targets.

The timing sequence of stimulus presentation within a single trial was as follows. The central fixation cross was displayed for 500 ms. Then the cross disappeared and, after a randomly variable interval between 133 and 1200 ms in which the background was black and empty, two bright square outlines (with the side subtending 4.4°) appeared in the upper left and upper right corners of the screen (with vertical and horizontal displacements subtending 3° and 8°, respectively) for a randomly variable interval between 733 and 1100 ms. This preparatory event was followed by the 67-ms display of a visual target (spiky/curvy shape in left/right position) in USV trials; by the 67-ms display of a visual target and the 200-ms presentation of the irrelevant sound (tzk/wow) in MSC and MSI trials; by the 70-ms presentation of the irrelevant sound only in catch trials.

3.8.3.2 Results

Average percentages of false alarms (responses to isolated sounds in catch trials) and misses, as well as RTs, were separately computed for left-right positions. Since a preliminary analysis showed that position had no effect on any dependent measure, I utilized 36 catch trials each to compute average false alarm percentages for tzk (1.8%) and wow (1.1%) sounds; such percentages were small and did not differ statistically, showing that participants were quite conservative in releasing key presses, independently of the specific sound.

As regards individual miss percentages and average RTs for hits, I computed them with reference to the 12-trial subset, irrespective of position, for each combination of the Condition (USV, MSC, MSI) $\times$ Contrast (1.0, 5.3, 20.7, 63.6%) $\times$ Shape (spiky, curvy) design. Figure 3.3 shows mean and s.e.m. values of miss percentages (upper graphs) and hit RTs (lower graphs). As expected, performance in multisensory trials was clearly better than performance in unisensory trials. Furthermore, both latency and accuracy data indicate that minimally salient targets (Michelson contrast= 1.0%) were harder to detect than targets with contrasts in the 5.3-63.6% range.

A 3-way ANOVA on the distribution of transformed miss proportions [x= arcsin $(pMiss)^{0.5}$] showed a main effect of Contrast ($F_{3,51}$= 10.25, p< 0.001), mainly attributable to the occurrence of more misses in trials with minimal contrast (25.2%) than in trials with the

three higher luminance contrasts (19.4% on average), and a main effect of Condition ($F_{2,34}$= 4.16, p< 0.05), consistent with the expected reduction of misses in multisensory (20.0%) over unisensory (22.6%) trials (MS vs. USV planned comparison: $F_{1,34}$= 5.49, p< 0.05). Miss percentages in multisensory congruent (20.2%) vs. incongruent (19.9%) trials did not differ and there were no significant effects involving target shape (main effect of Shape and 3-way interaction: F< 1; Shape × Condition interaction: $F_{2,34}$= 1.43, p= 0.25; Shape × Contrast interaction: $F_{3,51}$= 1.13, p= 0.34). The Condition × Contrast interaction was not significant (F< 1).

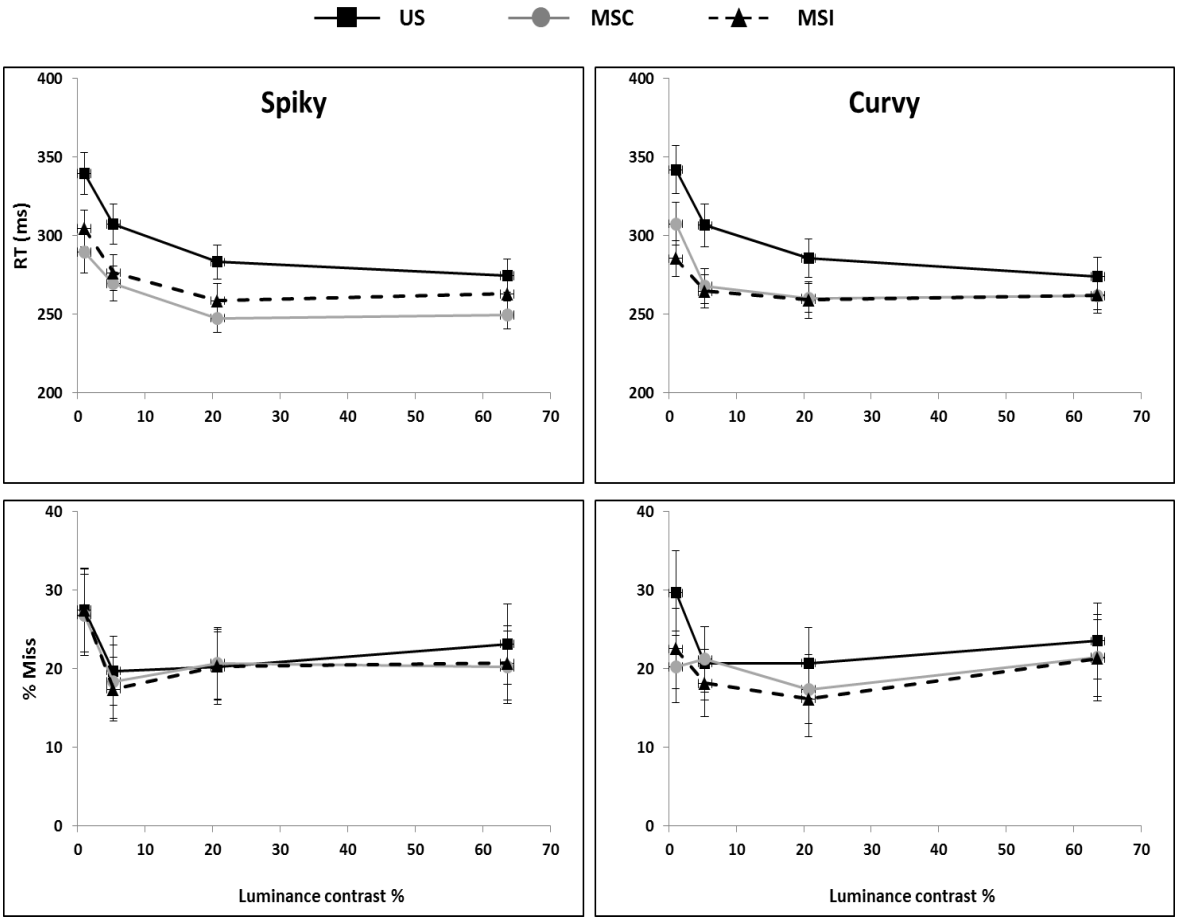


Figure 3.5 Experiment 5, % Miss and RT. Miss percentages (± 1 s.e.m., upper graphs) and RTs (± 1 s.e.m., lower graphs) for the Target (2) × Condition × Contrast (4) design of Experiment 5. Detection was clearly harder when the target luminance contrast was about 1%. In general, performance in multisensory congruent (MSC) and multisensory incongruent (MSI) trials was better than in unisensory trials.

A 3-way ANOVA on the distribution of transformed RTs (x= 1/RT) showed a main effect of Contrast ($F_{3,51}$= 59.83, p< 0.001), mostly dependent on slower responses in trials with minimal contrast (298 ms) relative to trials with the three higher luminance contrasts (261 ms on average), and a main effect of Condition ($F_{2,34}$= 41.07, p< 0.001), consistent with the expected RT reduction in multisensory (260 ms) over unisensory (290 ms) trials (MS vs. USV planned comparison: $F_{1,34}$= 81.60, p< 0.001). Average RTs in multisensory congruent (260 ms) vs. incongruent (261 ms) trials did not differ, but a congruency effect at higher contrasts was revealed by the following analysis on %MRE.

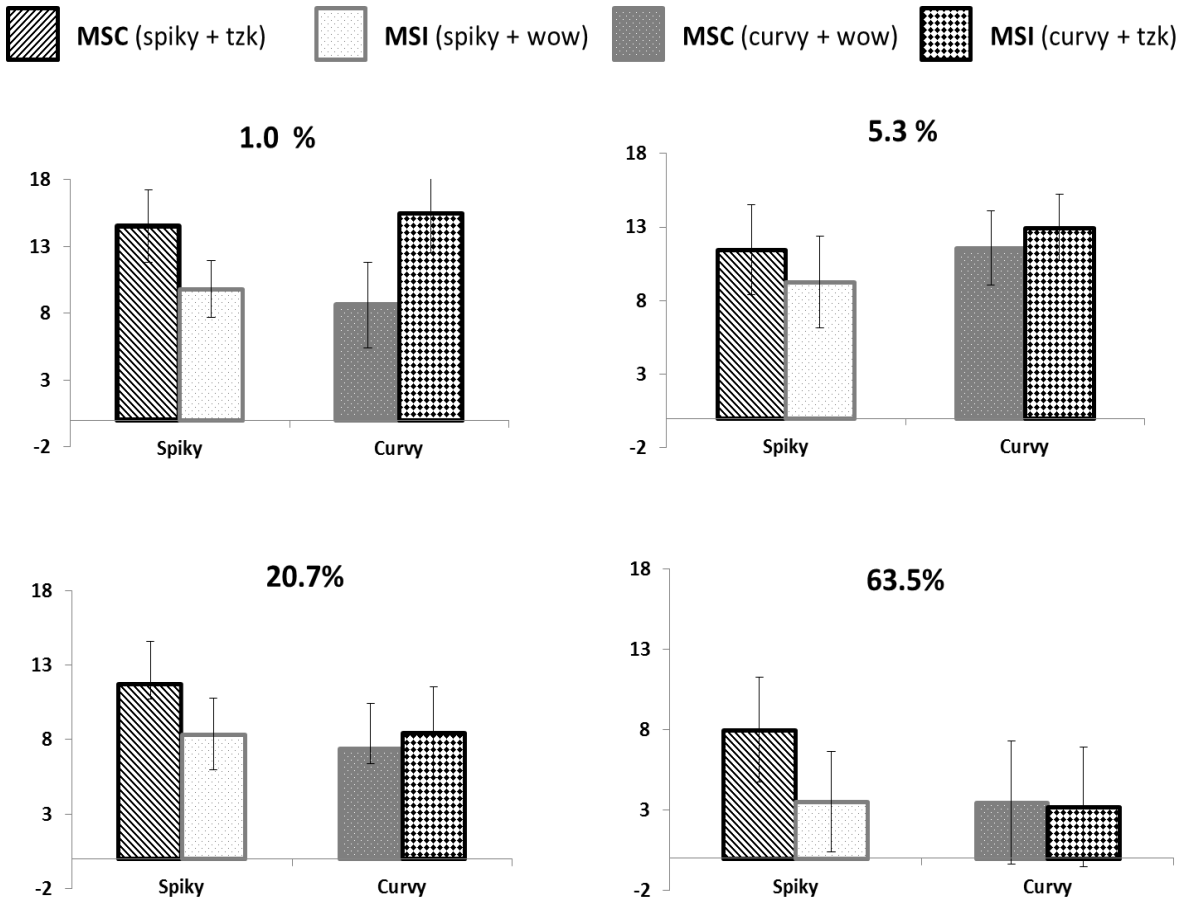


Figure 3.7 Experiment 5, % MRE (± 1 s.e.m.)**.** Multisensory response enhancement percentages (%MRE) in the 4 combinations of the Sound × Shape design, for each target luminance contrast.

With reference to RT values (or 1/RT transformed values in ANOVA) %MRE were defined as [100·(MSC-USV)/USV] and [100·(MSI-USV)/USV] for congruent and incongruent trials, respectively, and computed for every participant. Figure 3.7 shows the distribution of %MRE in the 4 sound-shape combinations, for each of the 4 target luminance contrasts. The Shape (Spiky, Curvy) x Contrast (1.0, 5.3, 20.7, 63.5%) x Congruency (MSC, MSI) within subject ANOVA revealed only a main effect of the Contrast ($F_{3,51}$= 3.55, p< 0.02) and an effect of the Congruency x Shape interaction ($F_{1,17}$= 9.64, p< 0.01), whereas there were no effects of Shape, Congruency or of their interactions (F< 1).

The results reported above indicate the presence of two effects. First of all, the main effect of the factor Contrast is in accordance to the predicted IER: an overall %MRE decreases as contrast increases. The inverse relationship between %MRE and target salience is displayed in Figure 3.8. The slopes of the straight line interpolating data for four contrasts are negative. To evaluate the statistical significance of the negative trends shown in Figure 3.8 I computed the individual slopes for every participant, in each of the 4 Shape x Congruency combination, with a total number of 50 negative slopes over 72 (one-tailed binomial test, p< 0.006).

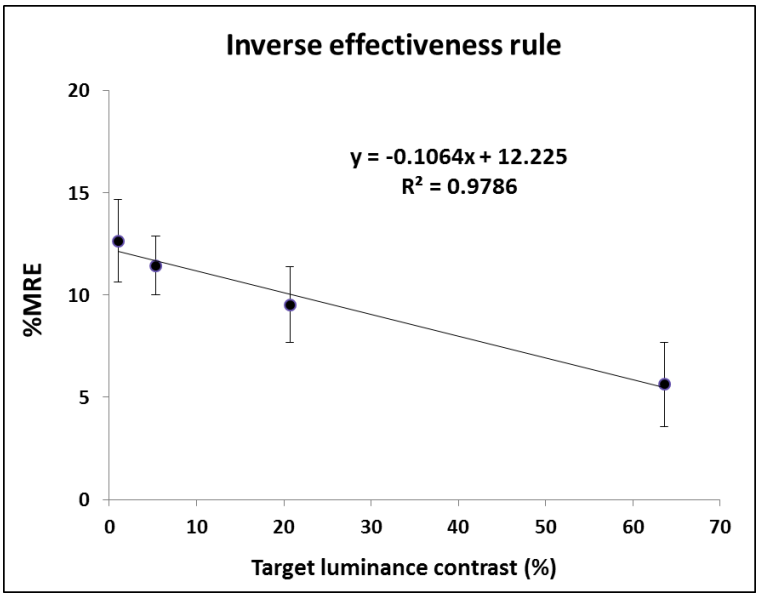


Figure 3.8 Experiment 5, Inverse effectiveness rule. Average % MREs (± 1 s.e.m.) for MS conditions plotted in relation to the target luminance contrast; the %MRE decreased as the luminance contrast of the visual target increased, consistently with the inverse effectiveness rule of MI.

The second effect regards the congruency effect. The interaction with the shape was due to the fact that, overall, a strong congruency effect was present with the spiky shape (MSC_{spiky} vs. MSI_{spiky}= 11.4 vs. 7.7; t(17)= 2.6 , p< 0.02) but not with the curvy shape (MSC_{curvy} vs. MSI_{curvy}= 7.8 vs. 10.0; t(17)= 2.0, p= 0.06, taking p< 0.025 as the significance level after the Bonferroni correction).

In Figure 3.9 the congruency gain for spiky and curvy target shapes is plotted against luminance contrast. The gain was defined as the difference between the %MRE for a congruent pairing (for instance, tzk + spiky shape) and the %MRE for an incongruent pairing (for instance, tzk + curvy shape).

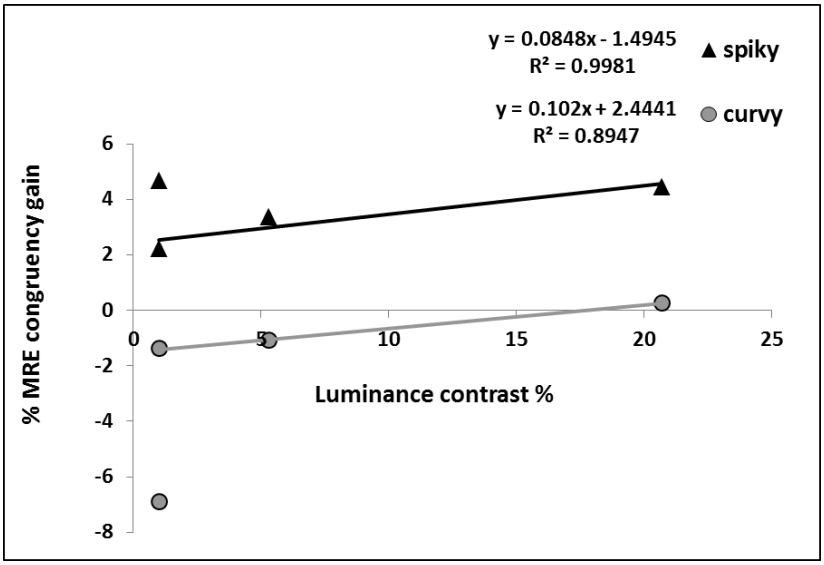


Figure 3.9 Experiment 5, Congruency effect. Congruency gain for the two shapes as a function of target contrast. The gain was defined as the difference between the two %MREs for the same visual shape: one relative to the pairing with the congruent sound and the other relative to the pairing with incongruent sound. Interestingly, the congruency gain increases as a function of target luminance contrast.

The minimal contrast was excluded from the interpolation because, as shown by miss percentages and RTs presented in Figure 3.5, detection of minimally salient targets was disproportionally difficult.

3.8.3.3 Conclusions

While I expected that the overall %MRE would follow the inverse effectiveness rule, on the basis of the general idea that the integration with an irrelevant signal in an unattended modality becomes less relevant when the visual target is more salient, the positive relation between the amount of the congruency gain and target saliency was unexpected. To say the least, one should conclude that the formal sound-shape congruency studied in Experiment 4 (i.e., the correspondence between the tzk sound and the spiky shape, as well as the one between the wow sound and the curvy shape) is qualitatively different from the spatio-temporal congruency supporting the integration of acoustic and visual stimuli into a unitary multisensory event. In the case of spatio-temporal congruency the %MRE is negatively correlated with target salience; while the formal sound-shape congruency gain is positively correlated with target salience. One might speculate that such a positive correlation follows from the analysis of complex intrinsic features of auditory and visual component events: only when the target is relatively salient such an analysis is possible, making their similarity effective and leading to a reasonable congruency gain.

As regards congruency, in Experiment 5 I replicated the effect described in Experiment 4 with spiky shape and sound, whereas no difference between the congruent and incongruent condition for the curvy shape and sound were observed. Admittedly, given the fact that the two pairings of stimuli were created ad hoc, we can assume that the curvy pairing was not adequate for the purpose of the experiment.

3.9 General conclusion

In Experiments 4 and 5 CMC were investigated in a simple detection task. A consistent congruency effect (faster responses in congruent vs. incongruent trials) were obtained when the spiky shape were paired with a spiky sound in both experiments, whereas no difference were evident for curvy shape paired with a soft sound. Even if the matching between spiky shape and spiky *tzk* sound and curvy shape and soft *wow* sound were performed prior to the experiment from a different group of participants, they had to select from a limited number of different choices. I would speculate that the group of stimuli were not suitably chosen for the purpose of my experiment, especially for what regards the curvy-soft combination, and therefore I did not obtain a superiority in that particular experimental condition.

Different aspects of the results from Experiments 4 and 5 need a discussion. The first main point is, perhaps, the automaticity of cross-modal interactions. In the experiment 4 and 5, participants performed the detection in a *focused attention task*, in which they had to respond to visual stimuli and ignore auditory stimuli. Moreover, the auditory stimuli were also part of the catch trial; that means, participants had also to inhibit responses to auditory stimuli. Therefore, independently from the direction of the selective attention, unattended stimuli are processed at a very early stage of the information processing, and this can tell us about the mechanism that underlies such interactions, suggesting an *early holistic processing*. According to this hypothesis, already proposed for intramodal interactions (Lockhead, 1972), *integral* cross-modal dimensions (such as lightness and pitch) are processes holistically, making it impossible to selectively attend to one separate modality. The non-separability of such dimensions is in accordance with the result coming from other studies (e.g. Parise & Spence, 2009; see introduction for the description). However, a further interpretation of the present hypothesis will be given in the general conclusion of the next chapter.

A second point regards the type of cross-modal interaction found in my experiments. The classification of cross-modal congruencies in statistical, structural and semantic leads supports the following speculations on the basis of my data. First of all, the interactions reported in my experiment could be described as *statistical*: it is easy to think

about many objects having similar shape and sound from our everyday living (e.g. the flashing light and the sound of the flash are similar to our spiky congruent condition). The brain could have abstract some general rules on the basis of the exposition to spiky objects producing spiky sounds. A second explanation can be semantically mediated: the word spiky is used (also in our explanation) both for the shape and sound. However, this is not so straightforward for the curvy shape-soft sound. In conclusion, the structural account seems to be less intuitive, but still valid. For example, we can think of the cross-modal congruencies related to dimensional changes (e.g. loudness and brightness) as being qualitatively different from those related to more complex property such as the shape. The first category refers to the *prothetic* dimension introduced by Stevens (1957) and regards all the continua that have to do with quantity or magnitude, e.g.: loudness, brightness, dimension etc. Moreover, Stevens introduced in this category of qualities also the *rate of change*. If we think of the curvilinearity as of a continuum, with the spiky and curvy being the two extremes of the continuum, it is easy to think about the stimuli from the present experiments as being similar to other types of dimensional congruencies. For example, we can abstract some information (e.g. number of interruptions) that defines our cross-modal stimuli, and that can be supposedly detected by a hypothetical *detector of discontinuities* present in our brain. Such a mechanism responsible for the detection of discontinuities bot in the temporal and spatial domain is well investigated in the literature (e.g. Schlaghecken, Meinecke and Schröger, 2001). In fact, discontinuities in the environment are an important source of information to the organisms. They help the organism to organize the sensory input into coherent perceptual objects and categorize stimuli into as figure vs. ground or target vs. non-target.

The effects related to the perception of discontinuities are present both in the auditory and visual domain. For example, a visual cluster may appear to “pop-out” from its background if its orientation is different from the orientation of the surrounding elements (e.g., Julesz, 1981; Nothdurft, 1991). Similarly, in audition we become aware of distortions from isochronous presentation of tones even when our attention is not oriented toward the auditory stimulus, suggesting that the mechanism may be pre-attentive (e.g., Näätänen, 1992; Schröger, 1997).

Another example of the a key role of the detection of discontinuities in the perception is represented by the well documented mismatch negativity effect, which regards a characteristic change in electroencephalographic and magneto encephalographic recordings, corresponding to a deviation of a sound or a visual feature among repetitive stimuli (see Garrido et al., 2009 for a review).

Given these assumptions, we can hypothesize that the CMC shape effect in my experiments may be the result of the congruency between the outputs of two detectors of discontinuities in the auditory and visual modality or, an alternative explanation may be the existence of a *cross-modal detector for discontinuities.* Such a hypothesis is purely speculative at this point of research, and additional experiments should clarify the mechanism involved in the cross-modal shape congruency effect.

CHAPTER 4

CROSS-MODAL SHAPE CONGRUENCY AND THE RACE MODEL ANALYSIS

Chapter 4:
Cross-modal shape congruency and the Race model analysis

Chapter overview. The previous chapters investigated various factors that affect MI such as the spectral components of stimuli (colour vs. luminance) and the spatial coincidence (Chapter 2) or the cross-modal shape similarities (Chapter 3). In both the previous chapters the tasks were performed in a focused attention task, suggesting that MI is automatic and pre-attentive, and that an auditory stimulus can affect the visual even when the attention is focused only to the visual modality. Moreover, both the inverse effectiveness rule and the congruency effect were obtained with stimuli of 4 different luminance contrasts. The pattern of results showed that the two effects had an opposite relation with the stimulus saliency: inverse relation in for the IER and a direct relation for the congruency effect.

In the Experiment 7 of the present chapter, the same stimuli from Chapter 3 were presented in a similar experimental design (unisensory, congruent and incongruent combination; spiky and curvy shapes, spiky -*tzk*- and soft –*wow*- sounds). Contrary to Chapter 3, in the present experiment the cross-modal shape perception has been investigated in a *redundant target task* in which participants should respond to both unisensory (either visual or auditory) and multisensory targets (either congruent or incongruent), and the attention is thus distributed cross-modally. Such a paradigm allowed performing the Race model analysis (Miller, 1982), which told us that both the congruent and incongruent pairs of stimuli were integrated (and thus, the mechanism underlying is the neural co-activation model). Moreover, the amount of violation was stronger for some congruent pairs of stimuli in comparison to incongruent pairs, suggesting a major relevance of the *interactive co-activation models* over the *independent co-activation model.*

4.1 Rationale of the experiment

The mechanism of interaction between two stimuli has been a matter of debate for many years. Beside the Race model (Miller, 1982; see Chapter 1), which states that the two stimuli are processed independently, alternative co-activation models have been proposed: e.g. the *Independent co-activation model* and *Interactive co-activation model*.

The *independent co-activation model* states that multisensory stimuli produce independent activations, and those activations are summed together to evoke the response (Morton, 1969). According to this model, the co-activation is the result of the adding process which is only dependent on the magnitude of the activation, and not by the identities of the targets that produced them. On the contrary, the *Interactive co-activation model* states that there can be an interaction between the processes detecting of two (or more) stimuli, and thus the response activation of a target in one channel is influenced by the identity of the target on the other channel. To link these models to the concept of CMC, we can hypothesise that according to the first model the responses in congruent and incongruent trials (e.g. experiments in Chapter 3) should be equally fast, whereas according to the interaction model, the congruency should be a facilitative factor in response co-activation.

This hypothesis was tested by Miller (1991). He presented auditory and visual stimuli in congruent and incongruent conditions, and asked participants to perform a speeded detection task to audio, visual and audio-visual pairs of stimuli. Different attributes of the stimuli were manipulated: the spatial position in the case of visual stimuli (upper, middle or lower position) and the pitch for auditory stimuli (high, medium or low), trying to replicate the mapping between the position and pitch found by other authors with a speeded classification paradigm (e.g. Marks, 1978). The authors reported that RTs to congruent pairs of stimuli were faster that RTs to incongruent. However, when the authors performed the analysis with multiple t-test comparisons (Miller, 1982) using the Inequality 1, no one of the violations were significant. The authors concluded suggesting that the Race model is an inadequate procedure for the testing of the hypothesis.

In the present chapter I tried to replicate the procedure adopted by Miller (1991), for the cross-modal shape congruency observed in Experiment 4 and 5.

4.1.1 Experiment 6

Experiment 6 was run to study the relationship between the congruency effect demonstrated in Experiments 4, 5 and MI, operationally defined by a superadditive MRE (i.e., by a violation of the Race Model Inequality, see Chapter 1). To test the Race Model Inequality I modified stimuli, apparatus and procedure of Experiments 4 and 5 in the following way:

- one luminance contrast (20.7%) was selected among the 4 utilized in Experiment 5, taking into account that the repetition of the same visual target should facilitate performance;
- Participants were required to distribute their attention cross-modally, responding to either auditory or visual events as well as to their pairing (redundant target paradigm), while not responding when the presentation of the two peripheral outline squares was not followed by any auditory or visual event.

This allowed us to compare RT distributions in unisensory (both auditory and visual) and multisensory (both congruent and incongruent) conditions, derive race model predictions and detect model violations supporting the possible occurrence of MI.

Even if the "curvy" stimulus combination did not result in any congruency effect, I decided to keep it in the design as a non-congruent pairing to compare with the congruent spiky couple.

4.1.1.1 Method

Participants

Eight healthy right-handed students (age range 19-25 years, 4 males and 4 females) took part in the experiment. All had normal hearing and normal or corrected-to-normal visual acuity. They were naïve to the purpose of the investigation, gave their informed consent prior to the beginning of the experiment, and received course credit for their participation

Apparatus and stimuli

I used the same apparatus and stimuli of Experiments 4 and 5, with the only exception of target luminance contrast, set at 20.7% in all trials involving visual targets.

Procedure

The experimental session included 4 subsets (unisensory auditory USA, unisensory visual USV, multisensory congruent MSC, multisensory incongruent MSI) of 80 positive trials each and one subset of 100 catch trials, as follows.

- USA trials: 20 for each combination of tzk/wow sound and left/right position, in the absence of a visual target;
- USV trials: 20 for each combination of spiky/curvy visual shape and left/right position, in the absence of an auditory target;
- MSC trials: 40 in which the tzk sound was paired with the spiky shape (20 for each spatial position) and 40 in which the wow sound was paired with the curvy shape (20 for each spatial position);
- MSI trials: 40 in which the tzk sound was paired with the curvy shape (20 for each spatial position) and 40 in which the wow sound was paired with the spiky shape (20 for each spatial position);
- Catch trials: 100 in which the presentation of empty square outlines in the upper left and right corners of the screen was not followed by a target, neither auditory nor visual.

Participants were instructed to press the left button of the mouse with the index finger of the dominant hand as quickly as possible to respond to the presentation of any visual target (either spiky or curvy, in either unisensory or multisensory conditions) and any auditory target (either tzk or wow, from the left or right loudspeaker). Moreover, they were asked to suppress any response when empty outline squares were presented without visual or auditory targets.

4.1.1.2 Results

Since sounds and/or shapes were presented randomly at the left or right side of the participant's visual field, I ran a preliminary analysis to test whether spatial position had any influence on dependent variables. No effect of spatial position emerged. Therefore, subsequent analyses were performed irrespective of it. The overall amount of false alarms was 1.6%, indicating that participants were generally accurate.

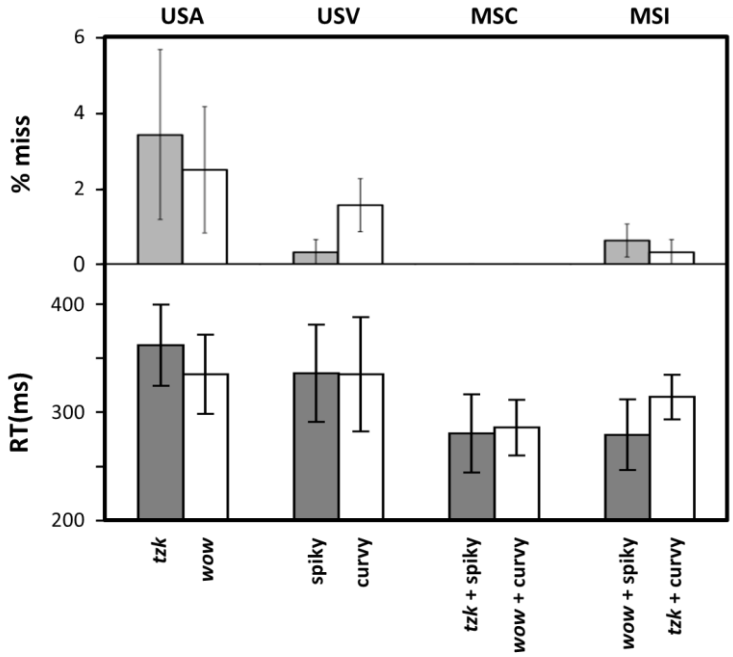


Figure 4.1 Experiment 6, % Miss and RT. Miss percentages (± 1 s.e.m.; upper graph) and RTs (± 1 s.e.m.; lower graph) in the 8 conditions of Experiment 6. Within unisensory conditions wow sounds were detected faster than tzk sounds. Detection was better in MS than US conditions (multisensoriality effect) and in congruent than incongruent conditions (congruency effect).

Figure 4.1 shows mean and s.e.m. values of miss percentages and RTs for hits in each of the 4 unisensory conditions (tzk and wow sounds, spiky and curvy shapes) and of the 4 multisensory conditions. A planned-comparison analysis was performed on transformed miss proportions [$x= \arcsin(pMiss)^{0.5}$] and response latencies ($x= 1/RT$), under the expectation that detection was better in multisensory than unisensory trials, as well as in congruent than incongruent trials.

First, consider unisensory trials. Miss percentages did not differ for either sounds (3.4 vs. 2.5% for tzk and wow sounds, respectively; t(7)= 1.39, two tailed, p= 0.21) or visual targets (0.3 vs. 1.6% for spiky and curvy shapes, respectively; t(7)= 2.00, two tailed, p= 0.09). However, wow sounds were detected faster than tzk sounds (334 vs. 361 ms; t(7)= 3.49, two tailed, p< 0.02); while RTs for spiky and curvy shapes did not differ (335 vs. 335 ms; t(7)= 0.41, two tailed, p= 0.69). Detection performance tended to be worst for sounds than visual shapes, but neither accuracy ($miss_{USA}$= 3.0% vs. $miss_{USV}$= 0.9%; t(7)= 0.77, two tailed, p= 0.47) nor response latencies (RT_{USA}= 348 ms vs. RT_{USV}= 335 ms; t(7)= 1.14, two tailed, p= 0.29) were significantly different.

Then, consider the overall superiority of multisensory stimulation. According to both dependent measures performance was better in MS than US trials. Average miss percentages were lower for multisensory events in both congruent ($miss_{USV}$ - $miss_{MSC}$= 0.9%; t(7)= 3.05, one tailed, p< 0.02) and incongruent ($miss_{USV}$ - $miss_{MSI}$= 0.4%; t(7)= 1.97, one tailed, p= 0.09) trials. On average, detection of multisensory events was faster than detection of the best unisensory stimulus in both congruent (RT_{USV} - RT_{MSC}= 53 ms; t(7)= 9.30, one tailed, p< 0.001) and incongruent (RT_{USV} - RT_{MSI}= 39 ms; t(7)= 5.86, one tailed, p< 0.001) trials.

Finally, consider the congruency effect. Performance was better in congruent than incongruent multisensory trials with respect to both accuracy ($miss_{MSI}$- $miss_{MSC}$= 0.5%; t (15)= 2.23, one tailed, p< 0.05) and response speed (RT_{MSC} - RT_{MSI}= 14 ms; t (15)= 2.51, one tailed, p< 0.05).

To evaluate the effects of multisensoriality and congruency, in Experiment 6 I derived from RT data a set of %MRE more specific than in Experiment 5. Since Experiment 6 provided reaction times for both auditory and visual targets, the %MRE was defined as the relative difference between the response speed in the congruent/incongruent multisensory condition and the faster response speed among the two, to either the auditory or visual target, for every participant, according to the following formula: MRE= $[100 \cdot (1/RT_{MS}\text{-}\max(1/RT_{USV},1/RT_{USA}) / \max(1/RT_{USV},1/RT_{USA})]$. Figure 4.2 shows the distribution of %MRE in the 4 sound-shape combinations. In this case I computed a 2-way ANOVA on %MREs using Congruency (MSC, MSI) and Shape (spiky, curvy) as main factors. In order to test the significance of the %MRE, four t test against 0 were run, indicating that

three contrast out of four were significant: MSC_{spiky}= 19.0 %, MSC_{curvy} = 12.0 %, MSI_{spiky} = 15.2 %, MSI_{curvy}= 7.4; t (7)= 7.05, p< 0.001; t (7)= 6.03, p< 0.001; t (7)= 10.9 , p< 0.001; t (7)= 2.48 , p< 0.04; critical p value after Bonferroni correction= 0.01). This analysis pointed out the fact that the improvement of the multisensoriality in the MSI_{curvy} condition (curvy shape + tzk sound) did not differ from 0. This point will be further explained at the end of the result exposition.

The ANOVA revealed a main effect of Shape ($F_{1,7}$= 8.48, *p*< 0.05) indicating that multisensory events involving the spiky shape (MRE_{spiky}= 17.13%) were more efficient than those involving the curvy shape (MRE_{curvy}= 9.7%) and a main effect of Congruency ($F_{1,7}$= 10.26, *p* < 0.02), with congruent conditions ($MRE_{congruent}$= 15.5%) more efficient than incongruent ($MRE_{incongruent}$= 11.3%) combinations. Interestingly, there was no Congruency x Shape interaction (F< 1), indicating that the congruency gain was the same for both shapes. This result is different from the one in Experiments 2 and 5, where consistently no effect of the curvy shape and soft sound congruency were found.

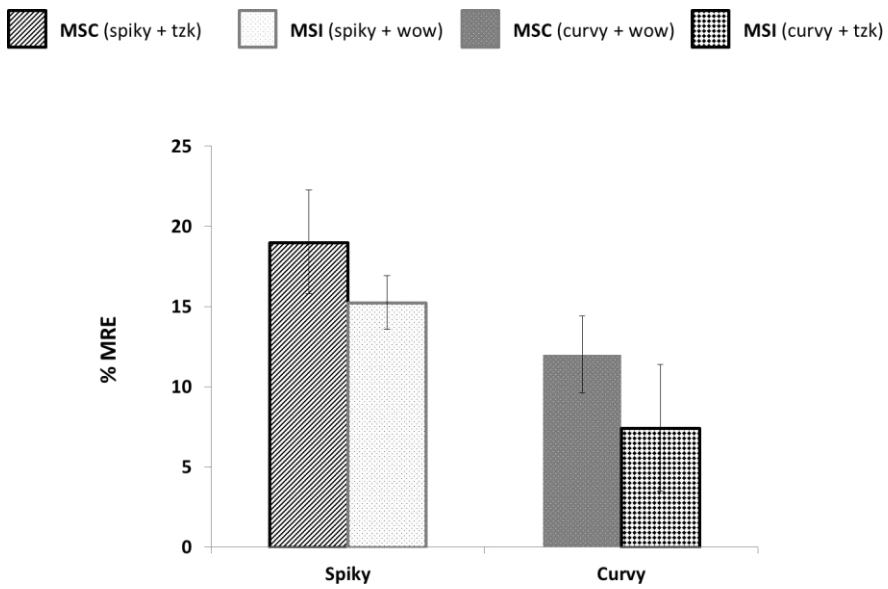


Figure 4.2 Experiment 6, % MRE [1]. Mean MRE percentages (± 1 s.e.m.), computed relative to the best US stimulus (see text), in the 4 sound-shape combinations. The congruency effect is clear, since the %MREs for [tzk + spiky shape] and [wow + curvy shape] congruent multisensory combinations were higher than the %MREs for [tzk + curvy shape] and [wow + spiky shape] multisensory combinations.

[1] Given the main effect of Shape, the graphical representation of congruent and incongruent stimuli in Fig. 4.3 is reversed in comparison to Fig. 4.2, in order to highlight the superiority of the congruent pairing in relation to the incongruent for each shape separately.

Race model analysis

Differently from Experiment 5 (run according to the focused attention paradigm) RT data obtained in Experiment 6 (run according to the redundant target paradigm) allowed me to perform Miller's RMI test to determine whether the observed superiority in multisensory trials could be explained by probability summation only or required a neural co-activation model (Miller, 1982; Ulrich, Miller, & Schroter, 2007).

The race model analysis (already described in the introduction chapter) assumes that the RT distribution for redundant stimuli is never larger than the sum of the RT distributions for the single stimuli. A violation of this inequality is interpreted as an indicator of an underlying neural summation (or co-activation) mechanism.

For every participant the 4 cumulative distributions of RTs were compared in multisensory conditions (one for each sound-shape combination) to the corresponding race model prediction derived from the sum of RT cumulative distributions obtained in unisensory conditions for the component auditory and visual stimuli.

Figure 4.3 shows the cumulative distributions of a typical participant for the 4 sound-shape combinations. In each of the 4 graphs the RT distribution for the multisensory conditions was on the left of the distributions for the two unisensory conditions (that were approximately superposed), consistently with a generic multisensoriality effect; furthermore, both MSC and MSI distributions were at least partially on the left of the race model line, evidence of a superadditive MI.

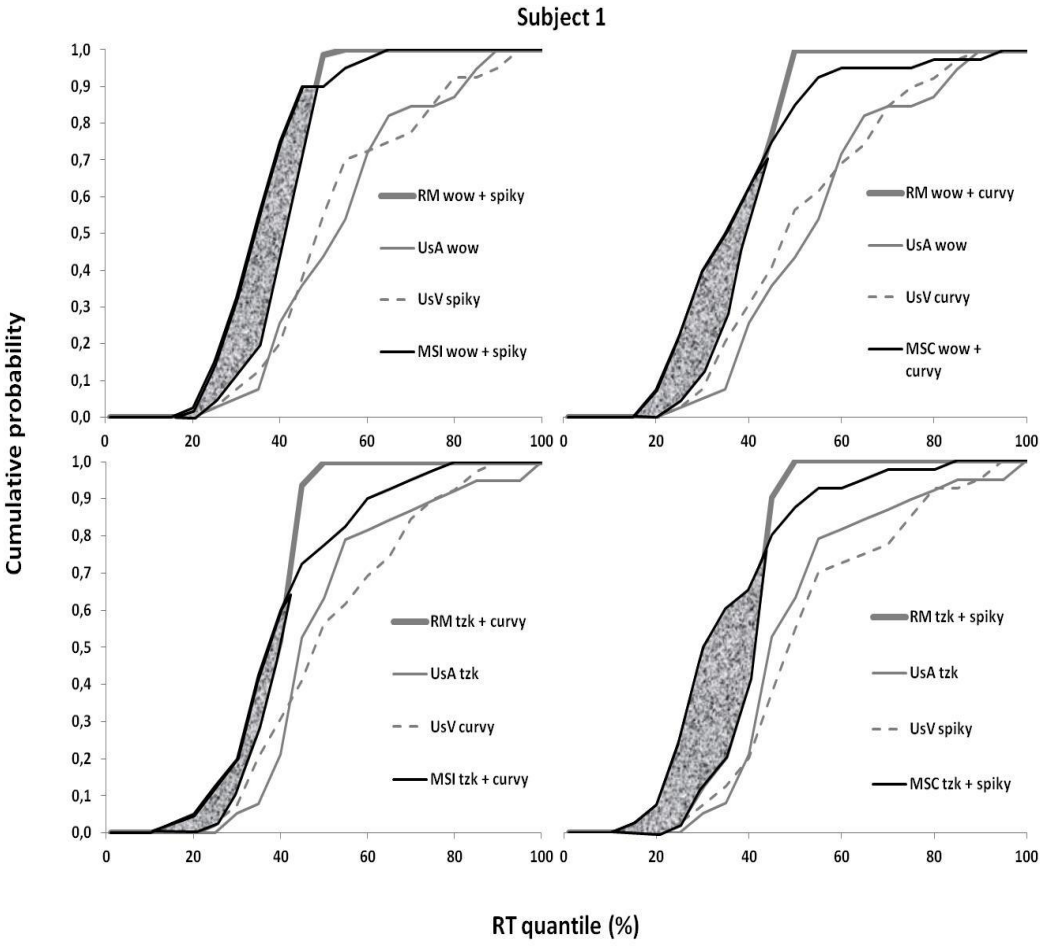


Figure 4.3 Experiment 6, Cumulative RT distribution functions in a representative subject. Cumulative RT distribution functions of participant S1, in the 4 sound-shape combinations. Each graph includes the distributions relative to tzk/wow alone [i.e., to the auditory stimulus in the unisensory condition], to the spiky/curvy shape alone [i.e., to the auditory stimulus in the unisensory condition (USV)], to either the congruent (MSC) or incongruent (MSI) multisensory event, as well as the distribution predicted by the race model [i.e., the sum of the cumulative distributions relative to the auditory and visual component stimuli]. Textured regions represent race model violations.

In graphs like those in Figure 4.3, the amount of RM violation was defined as the area of the textured region on the left of the race model line, bounded by the multisensory line (either MSC or MSI).

Mean RM violation areas (in arbitrary units) and s.e.m. are plotted in Figure 4.4. A Congruency (2) × Shape (2) ANOVA on violations area (in arbitrary units) revealed a main effect of the factor Congruency ($F_{1,7}$= 14.01, p< 0.01), and of the factor Shape ($F_{1,7}$= 13.4, p< 0.01). The effect of the Shape was due to the fact that overall the violation area of the

Spiky shape was larger than the violation area for the Curvy shape (6.66 vs. 3.81 respectively).

The Congruency x Shape interaction was also significant ($F_{1,7}$= 9.15, p< 0.02). Note that the congruent [wow + curvy] combination led to a RM violation much larger than the incongruent [tzk + curvy] combination (mean areas= 5.8 vs. 1.8); while the congruent and incongruent combinations involving the spiky shape did not differ significantly (area[tzk + spiky]= 6.33 vs. area[wow + spiky]= 6.99).

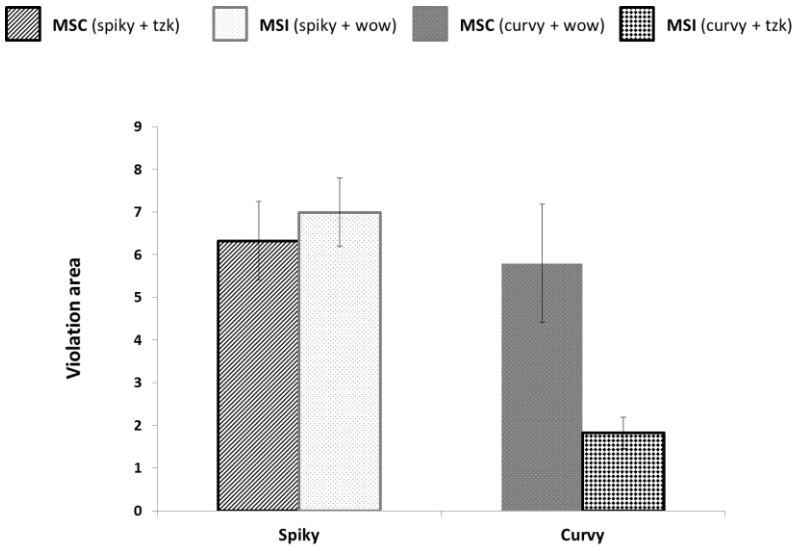


Figure 4.4 Experiment 6, Race model violation areas. Mean values (± 1 s.e.m.) of race model violation areas for the group of 8 participants, for each of the 4 sound-shape combinations. The pattern of violations suggest that the performance improvement in multisensory trials was always superadditive; and that it was larger in congruent curvy vs. incongruent curvy combination.

Mean RM violations shown in Figure 4.3 should not mask the fact that I also obtained some evidence of response depression for one of the two incongruent multisensory combinations. Figure 4.5 shows the cumulative RT distributions provided by 4 participants in the incongruent [tzk + curvy] condition. A weak RM violation was evident only in the leftward portion of the graph for three participants and no RM violation at all for participant S4. In the rightward part of the graph, relative to slower RTs, the distributions relative to the MS condition fall on the right of the distribution relative to the more efficient unisensory stimulus (and in certain cases of both unisensory distributions).

At least in some trials and for some participants, the sound-shape formal incongruence of the multisensory target deteriorated detection performance, relative to detection of component unisensory stimuli.

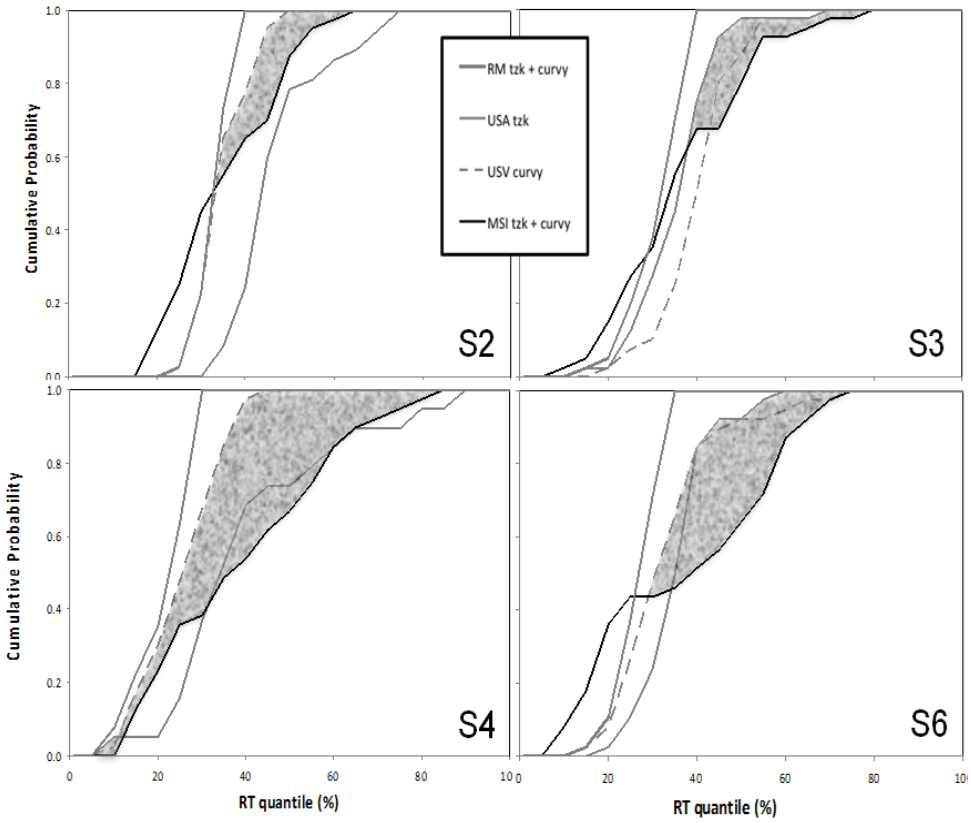


Figure 4.5 Experiment 6, multisensory depression. Cumulative RT distribution functions of participants S2, S3, S4, and S6, in the incongruent [tzk + curvy] combination. The textured regions mark cases in which RTs in the multisensory condition were slower than RTs for the most efficient (often to the least efficient too) unisensory stimulus, and can be interpreted as evidence of response depression.

In order to test the "profile of violation", RT corresponding to every percentile separately has been derived for the MSC and MSI (both spiky and curvy) conditions, and for the 4 different Race models for every subject separately. For every percentile I calculated the difference between RTs of each RM and the correspondent MS and

obtained a difference profile of the violation in relation to the RM. A series of t tests against 0 has been calculated on RT differences for every percentile to test the difference between the two AV conditions and the two models (Figure 4.6). For a step-by-step explanation see the APPENDIX of the present thesis.

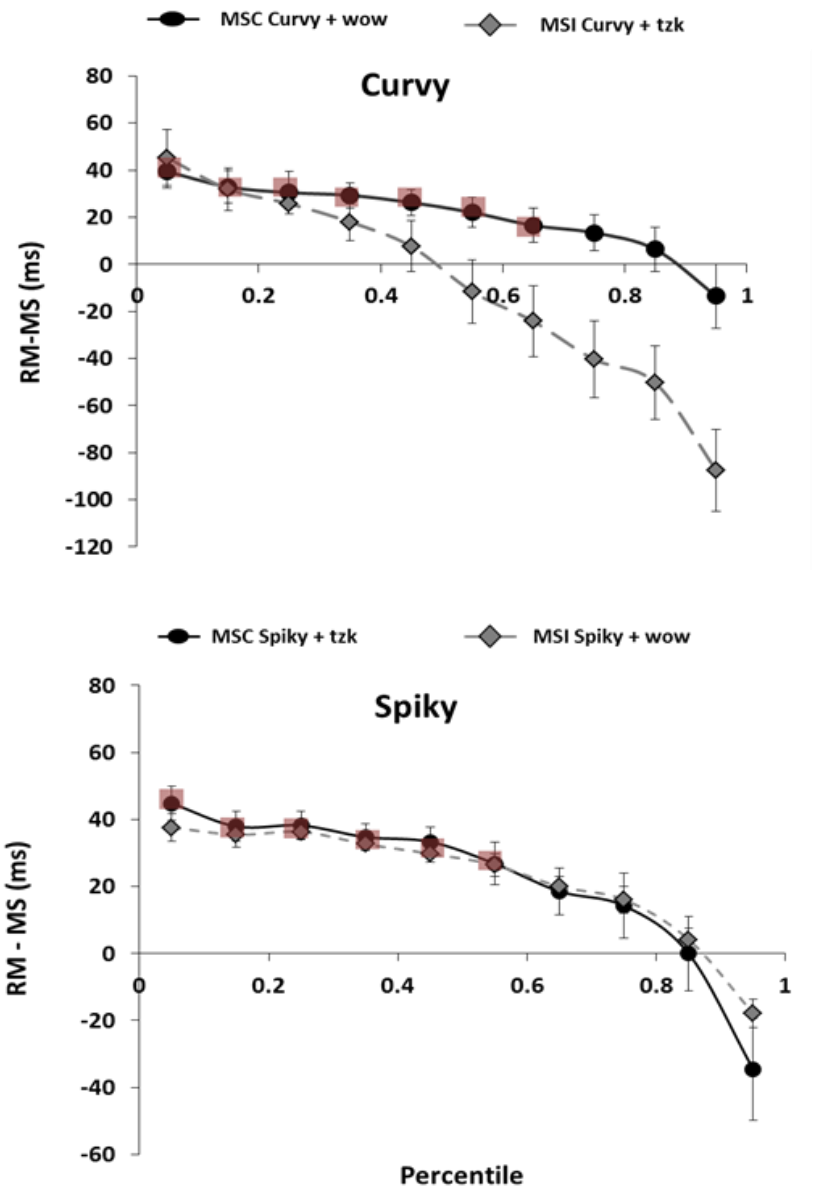


Figure 4.6 Experiment 6, Group difference profile (± 1 s.e.m.) for the four MS conditions (MSC_{curvy}, MSI_{curvy}, MSC_{spiky}, MSI_{spiky}) in relation to the corresponding Race models for every percentile. Red squares indicate the percentiles that violated the Race model. The MSC_{curvy} condition exhibited a strong violation of the model in 6 percentiles, while the MSI_{curvy} condition violated the model only in the first 2 percentiles. The spiky condition exhibited a similar strong violation both for MSC and MSI condition.

4.2 General conclusion

In the present Chapter 4, the Race model analysis has been used to test the cross-modal shape congruency effect described in Chapter 3. Contrary to Experiments 4 and 5 that were performed in a focused attention paradigm, in Experiment 6 participants performed the task in a redundant target paradigm, in which they had to distribute their attention cross-modally and respond to visual, auditory and multisensory events. The main difference obtained in Experiment 6 was the congruency effect with the curvy shape and soft sound, which I failed to elicit in Experiments 4 and 5. Supposedly, the distribution of the attention across modalities facilitated other less automatic congruencies (such as that between the curvy shape and soft sound) to appear.

The more important contribution, perhaps, is that congruency effects are observed in a simple detection task in which no discrimination is needed, in both focused attention (Experiments 4 and 5) and redundant target tasks (Experiment 6). Contrary to speeded classification, usually adopted in the study of the CMC, speeded detection is consistent with the assumption that such interactions occur at an early level of information processing.

Secondly, even if I did not have any prediction, I observed a congruency effect in accuracy. Even if the overall amount of the facilitation is almost negligible, the difference reached significance, proving that congruent pairings of stimuli facilitate the performance both in the speed of the response and in the accuracy.

Moreover, the Race model analysis was performed, showing that both congruent and incongruent events violated the model. The area of violation permitted to compare the strength of violation among different conditions. Both MSC and MSI for the spiky shape condition violated the model. However, the violation for the MSI_{curvy} was weaker in comparison to the other conditions (Figure 4.4 and Figure 4.6). If we look carefully at the data, we can find some indications of a depression (Figure 4.5): in some cases, the response in the MSI_{curvy} condition was worst that one or both unisensory condition. Unfortunately, the fact that we can observe such a clear pattern only in 4 participants out

of 8 does not allow us to make any strong conclusions out of the data. However, if we concentrate only on the %MRE analysis, we can see that no significant enhancement was observed in the MSI_{curvy} condition in relation to the best US response: on the basis of this we can assume that the incongruency between curvy shape and tzk sound did not allow for a superadditive MI.

To sum up, Experiments 4-6 indicate that the brain can abstract information about the curvilinearity-spikiness of visual contours and sounds, and facilitate the response to congruent vs. incongruent pairings. This effect is stronger when attention is distributed cross-modally, but can arise (only for the spiky combinations in the case of my experiments) also when attention is focused on the visual modality. A further discussion on the data will be provided in the General discussion chapter.

CHAPTER 5

AUDIO-VISUAL INTERACTION IN GENERATION OF SACCADES

Chapter 5
Audio-visual interactions in generation of saccades

Chapter overview. In Chapter 1, 2 and 3 the MRE was studied in manual responses, in which participants should respond by pressing a button in a simple detection task. In the present Chapter, a fundamentally different paradigm and approach is adopted. Since the neural structure most commonly linked with MI is the superior colliculus (SC), which is also the principal structure involved in saccade generation, the present chapter is dedicated to saccadic responses, adopting the *Saccadic inhibition paradigm* (SI; Reingold & Stampe, 2002). The saccadic inhibition concerns the inhibiting influence of a visual transient which result in a characteristic dip in the distribution of saccade frequency. Four experiments were run with the aim of testing a multisensory nature of the inhibiting interactions in the saccadic system.

5.1 Multisensory integration in generation of saccades

Saccades are rapid and ballistic eye movements, with the general goal to bring the object of interest to the fovea. They can be triggered by specific events in the environment or can be elicited by the voluntary effort to move the eyes without a specific target (Cassin & Solomon, 1990). When triggered by an event, the saccadic system has to select a target, which is fundamental to isolate just one target from a vast number of competing targets. Although saccades are usually investigated in a context of visual perception, they can be triggered by auditory and tactile stimuli, so target selection has to operate in a multisensory context (Blanke & Grüsser, 2001; Zambardieri, 2002).

In natural surroundings, the information that stimulates different sensory modalities, and triggers a saccade, occurs in various combinations: for example, two stimuli (e.g. audio and visual) may originate from the same object, and constitute a multisensory target for the saccadic movement. In this case, the target selection will be facilitated by the

co-occurrence of two stimuli, resulting in a redundant target effect (RTE; i.e. reduced RT, see Chapter 1; Todd, 1912). The strength of the facilitation will depend on the spatial and temporal configuration of the stimuli, and will be strongest when they overlap spatially and temporally (Colonius & Arndt, 2001; Corneil, Van Warnooij, Munoz & Van Opstal, 2002; Engelken & Stevens, 1989; Frens, Van Opstal & Van der Willigen 1995, Harrington & Peck, 1998; Hughes, Reuter- Lorenz, Nozawa & Friederich, 1994). If the spatial or temporal distance between the stimuli is increased beyond some critical point, they will be perceived as originating from different objects. As such, the two stimuli will compete as possible targets of the saccade and the competition will result in a response cost, either slowing down the SRT or changing the kinematic of the saccade. In such a case, only one of the two stimuli will be the target of the saccade, while the second stimulus will become a distractor. This phenomenon, known as the Remote distractor effect (RDE), has been well described in the case of visual targets and distractors (Levy-Schoen, 1969; Findlay & Walker, 1999).

The RTE and the RDE have been studied for years in a parallel way, in an attempt to understand the rules that govern the two effects. While the first has been described both intra-modally and cross-modally, the second has been investigated mostly in the visual domain. It is essential to understand the relation among the two effects in order to put more light on the mechanism of the target selection in a multisensory environment. Researchers should, therefore, adopt a multimodal approach as a necessary step forward in the study of saccadic eye movements. In the present chapter I will describea first attempt to study these saccadic phenomena within a multisensory integration paradigm.

5.1.1 Neurophysiology of the saccade system

Figure 5.1 illustrates the principal brain structures involved in the generation of saccades (from Reingold & Stampe, 2000). Before a saccade is generated toward a visual stimulus, a visual signal is transmitted from the retina to cortical and subcortical visual pathways. The cortical pathway projects to the Lateral Geniculate Nucleus (LGN) of the thalamus and the LGN projects almost exclusively to the striate visual cortex. The

subcortical pathway sends inputs from the optic tract to the superficial layers of the SC and represents the main route in the generation of saccades. Some of the characteristics of the SC relevant to multisensory integration have been described in Chapter 1. In the following material, the SC will be considered more in relation to its saccade generation role.

The SC is crucial both for visual fixation and the generation of saccadic gaze shifts (Corneil, Olivier, & Munoz, 2002; Freedman and Sparks, 1997; Munoz, Dorris, Parè & Everling, 2000).

The SC supports fast oculomotor responses to visual (or other modality) targets by directly activating the saccade generator in the brainstem. Its integrity is crucial for the production of short-latency reflexive saccades, including the shortest latency "express" saccades with a peak around 100 ms (Fischer & Ramsperger, 1984; Sparks, Rohrer & Zhang, 2000). Moreover, the SC receives top-down projections from many cortical areas that control both visual attention and saccades: the saccade-related activity of the SC neurons is shaped by inputs from the posterior parietal cortex, the frontal eye fields (FEF), and subcortical areas such as the substantia nigra pars reticulate (Hikosaka & Wurtz, 1985; Hanes & Wurtz, 2001; Wurtz, Sommer, Paré & Ferraina, 2001). The SC sends projections to the cortex through the internal medullary lamina (IML) of the thalamus, which play an important role in spatial shifts of visual attention as well as saccades (Figure 5.1; Reingold & Stampe, 2000; McHaffie, Kruger, Clemo& Stein, 1988).

The superficial and intermediate layers of the SC receive projections from cortical regions involved in saccade planning and visual attention: the FEF (Segraves & Goldberg, 1987), supplementary eyes fields (SEF, Shook, Schlag-Rey, & Schlag, 1990), and lateral intraparietal area (LIP; Lynch, Graybiel, & Lobeck, 1985). In the prefrontal cortex, the activity of FEF neurons reflects the selection of the visual target for a saccadic eye movement when several potential goals for movements are available. The role of the FEF is also to suppress reflexive saccades and to generate saccades to targets from modalities other that visual. The SEF is the second prefrontal area involved in this circuit: it receives inputs from visual areas similar to those of the FEF and sends projections to the saccade generator and SC (Shook et al., 1990). The LIP area in the parietal cortex is a structure that usually responds to the onset of a target and encodes a visual saliency map (Gottlieb, Kusunok & Goldberg, 1998). This response is particularly enhanced when the target is

attended covertly (Colby, Duhamel & Goldberg, 1996), suggesting that the neurons are activated preferentially by targets that have a high behavioural significance.

Different types of neurons important for the execution of the saccades are present in the SC. *Saccadic neurons* increase their discharge before and during the saccade, and are distributed in the intermediate SC. *Fixation neurons* are active during visual fixation and silent during the execution of saccade, and are present mainly in the rostral pole of the SC, where the fovea is represented. Their reciprocal activity is determined by inhibitory lateral connections (Meredith & Ramoa, 1998; Munoz & Istvan, 1998). The way in which the MI operates at this level of lateral connections has not been fully unveiled so far. However, sensory inputs from different modalities converge in the intermediate layers of the SC and selectively activate either fixation or saccades neurons.

The saccades neurons can be further divided in *build-up* and *burst* neurons (Munoz & Wurtz, 1995a). During a saccade, the activity starts in the build-up neurons, and spreads rapidly to the front of the SC, after which the build-up and burst neurons both discharge. The saccades and fixation neurons project to the reticular formation (Gandhi & Keller, 1997), where the *brainstem saccade burst generator* (BSBG) is located (Moschovakis, Scudder & Highstein, 1996; Corneil, Olivier et al., 2002). The principal role of this structure is to transform saccade commands into a precisely defined burst of activity from the oculomotor neurons, contracting the ocular muscles (Robinson, 1970). Both the direction of the saccade (King, Lisberger & Fuchs, 1986) and its amplitude (Hepp & Henn, 1985) are determined by the activity of the BSBG. The first is regulated by the relative activation of each of the six ocular muscles while the second is regulated by the total number of spikes in the saccadic burst. The BSBG consists of long-lead burst neurons (LLBN), medium-lead burst neurons (MLBN) and omnipause neurons (OPN). The OPNs act as an inhibitory gate on the BSBG (Munoz & Wurtz, 1993).

The rostral pole of the deep layers of the SC (DLSC), which corresponds to the central visual field, excites the group of neurons above mentioned, controlling in this way the active fixation and inhibition the saccade generation.

The deep SC receives extensive motor, premotor and multimodal sensory projections, and projects to the thalamus and the brainstem (Sparks & Hartwich-Young, 1989). The characteristics of neuronal activity in the deep SC have been broadly explained

in Chapter 1. The involvement of these neurons in the generation of saccades from multisensory sources will be described in the next paragraph.

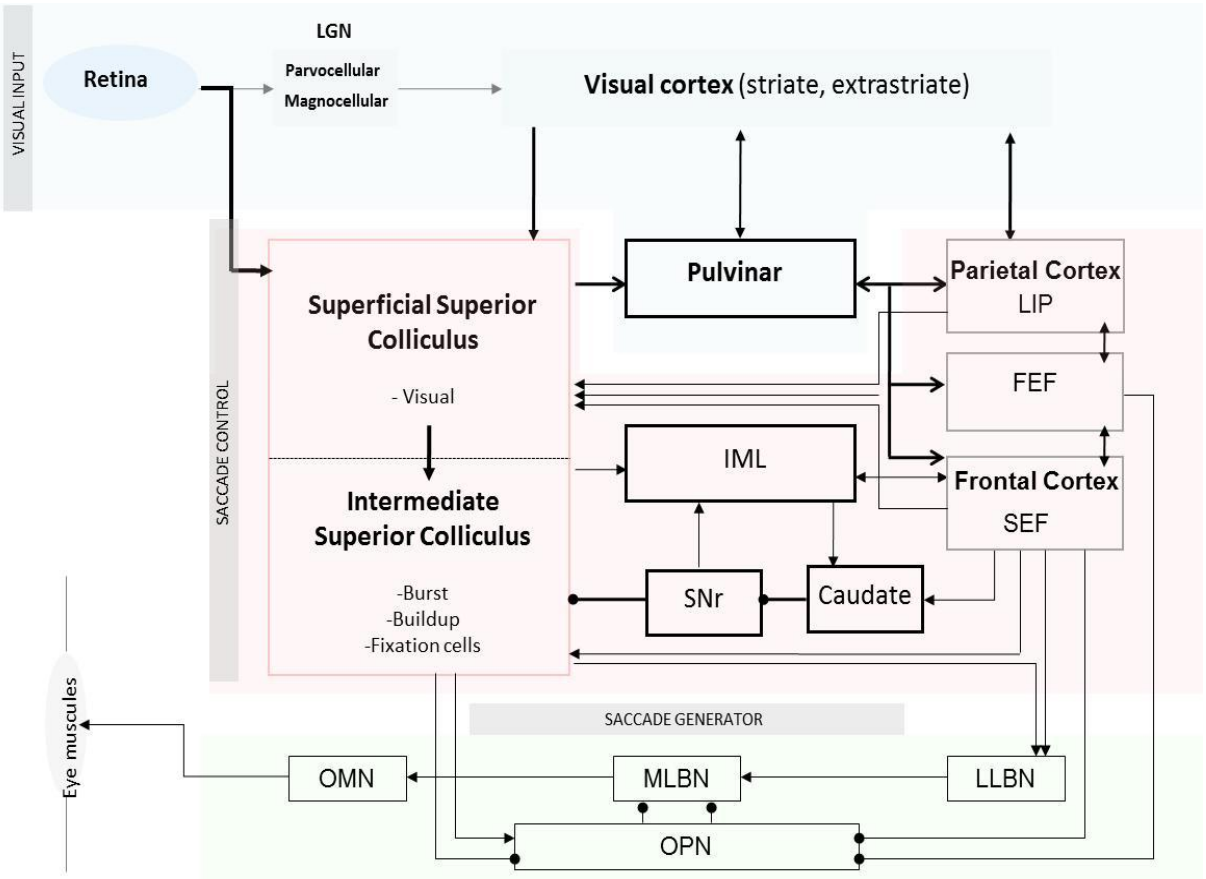


Figure 5.1. A schematic representation of structures involved in saccade generation. *From Reingold and Stampe (2000)*.The input proceeds clockwise from the retina to the eye muscles, through structures involved in visual input, saccade control, and the saccade generator. The neural connections are represented by the arrows, which are excitatory and/or inhibitory. Abbreviations: FEF-frontal eye field; IML-internal medullary lamina of the thalamus; LGN-lateral geniculate nucleus; LIP-lateral intraparietal area; SEF-supplementary eye fields; SNr- substantia nigra pars reticulate, LLBN-long-lead burst neurons, MLBN- medium-lead burst neurons, OPN- omnipause neurons, OMN- oculomotor neurons.

5.1.2 Neurophysiological basis and models of multisensory interactions in saccade generation

The SC contains some neurons that are completely dedicated to sensory integration, while others are involved in the initiation and control of saccades in goal oriented behaviour (Sparks & Hartwich-Young, 1989; Munoz & Fecteau, 2002, for a review). Goal oriented behaviours, such as moving the eyes toward, or reaching for a target (visual or auditory) require the system to compute the location of events in different reference frames. When the goal of behaviour is a combination of audio-visual stimuli, the two different sensory organs (eyes and ears) code their respective information in different ways: visual information is initially coded *retinocentricaly*, while the auditory information is coded *craniocentricaly* (in relation to the pinnae, which are immobile with respect to the head). In order to generate a response to audio-visual stimuli (or any other multimodal stimulus) information from the different modalities must, at some point, be represented within a common frame of reference. In the SC, each of the different sensory modalities appears to be represented in the form of a topographical "map" of the retina, of the auditory space or somatosensory map of the body. The co-registration of the different sensory maps is performed by multisensory neurons whose receptive fields are in register with one another (Meredith & Stein 1986; for a review, see Stein & Meredith, 1993). The system must transform the representation of a sensory stimulus into one that is appropriate for the motor act (the eye movement in the case of this chapter). Moreover, a neural processes, presumably in the posterior parietal cortex, that transform auditory and visual spatial coordinates into a common, trunk-centered, frame of reference was also hypothesised by other authors (Lewald, Dörrscheidt & Ehrenstein, 2000).

Based on electrophysiological data, Frens et al. (1995) proposed a neurophysiologically plausible model of how visual and auditory stimuli interact in triggering saccades (Figure 5.2). Saccadic eye movements are encoded by a large population of cells in a two-dimensional motor map in the FEF (Bruce & Goldberg, 1985) and in the DLSC (McIlwain, 1982; Sparks, Holland & Guthrie, 1976). Frens et al (1995) proposed that when a visual and an auditory target are spatially close, the populations that

would generally encode for a single target position might merge into a single region of activity. This region of activity will have a peak activity at the region that encodes the average position between the auditory and visual region.

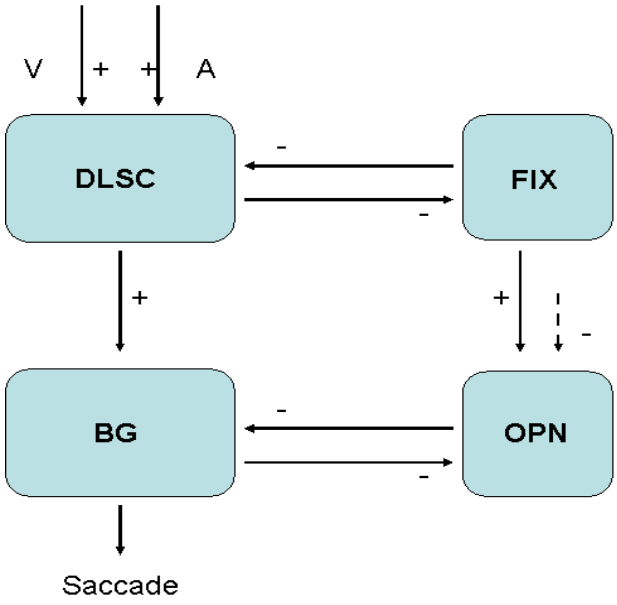


Figure 5.2 Model of audio-visual interactions in saccade generation (*from Frens et al., 1995; Munoz & Wurtz, 1993*). The structures that are involved are the following: DLSC (deep layers of the SC); BG (brainstem saccade burst generator); FIX (fixation neurons); OPN (Omnipause neurons). Visual (V) and auditory (A) signals project to saccade-related burst neurons in the DLSC. Bimodal interactions may either facilitate or detain the crossing of a threshold, necessary to silence the FIX. These neurons act through the OPN as an inhibitory gate on the generation of a saccade by the BG.

The *omnipause neurons* in the brainstem act as an inhibitory gate on the BSBG (Munoz & Wurtz, 1993). These groups of neurons are thought to be excited by the fixation neurons in the rostral part of the DLSC, maintaining active fixation by inhibiting saccade generation. In order to produce a saccade, a certain threshold of activity needs to overcome the fixation activity. Multisensory facilitation operates at this level; by enhancing the firing rates of pre-saccadic neurons of the SC, the crossing of this threshold is facilitated, and the saccade latencies therefore reduced.

5.1.3 Properties of visually and auditorily-driven saccades

In the present chapter, visually and auditorily guided saccades will be examined in the same experimental design. For this reason, I will provide a brief review of unimodal visual and auditory saccade response characteristics.

In general, the saccade latency onset is shorter for auditory targets in comparison to visual targets, regardless of target position in space (221 ms for auditory targets and 262 ms for visual targets; Zambarbieri et al., 1982). By contrast, the latency of auditorily-driven saccades is related to target position with respect to the eyes is generally slower in comparison with visually-driven saccades (Zahn et al., 1978; Zambardieri et al., 1982). For the same saccade amplitude, auditorily-driven saccades have a longer duration (i.e. the time during which eye velocity stays greater than a given threshold value) and a lower peak of velocity. Zambardieri et al. (1982) reports that a 20°auditorily-driven saccade has the duration of 92 ms, while a visually-driven saccade of the same amplitude has a duration of 77 ms. In addition, the latency of auditorily-driven saccades is reduced when the eccentricity increases from the head midline up to 40° of eccentricity (Zambardieri et al., 1982), since the auditory localization is less certain for targets at small eccentricities. In contrast, visually-driven saccades latency increases with increasing eccentricity (Kalesnykas & Hallett, 1994). Finally, auditorily-driven saccades are less accurate, with a 3° of average final error (Zambardieri et al., 1982), although Yao and Peck (1997), considering only single saccade responses, found that for target eccentricity of 10°, the accuracy of visual and auditory saccades were not significantly different. It must be noted, however, that all the differences depend strictly on the intensity of the stimuli, the properties of the sound source and the eccentricity of the stimulus, and should not be an obstacle for the MI when two different stimuli are presented as possible targets of the saccade (e.g. with a time-window of integration; see next paragraphs).

5.1.4 The influence of multisensory integration on saccade generation

A well-known effect is that saccadic reaction time (SRT) is reduced for spatially and/or temporally co-occurring visual and auditory stimuli, with respect to unimodal audio and visual stimuli (Colonius & Arndt, 2001; Corneil, Van Warnooij, Munoz & Van Opstal, 2002; Engelken & Stevens, 1989; Frens, Van Opstal & Van der Willigen 1995, Harrington & Peck, 1998; Hughes, Reuter- Lorenz, Nozawa & Friederich, 1994) or somatosensory stimuli (cf. Groh & Sparks, 1996a, for monkeys; Amlot, Walker, Driver, & Spence, 2003; Diederich, Colonius, Bockhorst, & Tabeling, 2003, for humans), resulting in a MRE, similarly to manual responses. Different factors may reduce SRT. According to the Race model (Rab, 1962), observers could respond simply to the onset of the auditory or the visual stimulus separately, and the reduction of the SRT could be due to simple statistical facilitation (see Chapter 1 for explanation of the Race model; Miller, 1982).

The Race model inequality tests (Miller, 1982), originally developed for manual reaction times, have been adopted in the analysis of MI in the generation of saccades. Similarly to manual RT, a common result coming from this area is that MRE in SRT to spatially coincident bimodal stimuli cannot be explained by statistical facilitation alone but rather by a neural co-activation mechanism (Corneil & Munoz, 1996; Diederich&Colonius, 2004; Harrington & Peck, 1998; Hughes et al., 1994; Arndt & Colonius, 2003), thus proving that the stimuli are not processed independently but that they co-activate the saccadic response synergistically.

The MRE in saccade generation is modulated by changes in sensory properties of unisensory stimuli. When the intensity of the target is decreased, the size of the MRE increases (Colonius & Arndt, 2001; Corneil et al., 2002), reflecting the *inverse effectiveness rule* observed in neurophysiological studies (Stein & Meredith, 1993; Perrault et al., 2005) and behavioural studies with manual responses (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003).Visual and auditory stimuli can also combine to influence saccade trajectories. Lueck et al. (1990) examined saccades to simultaneous of visual and auditory stimuli at increasing spatial disparities. They found that saccades to visual targets with a close, but spatially disparate auditory co-stimulus

produced saccade amplitudes with end points that represented an 'averaging' of the two saccade target positions, reflecting a phenomenon first described in visually-elicited saccades as the 'global' effect (Walker et al., 1997; Edelman et al., 2007; Findlay, 1982; He & Kowler, 1989). However, the influence of the MI on the spatial properties of the saccades is beyond the scope of my thesis.

Beside the inverse effectiveness rule, the saccadic responses follow also the spatial and temporal rule of the MI, which states that the integration among the stimuli is stronger when they are spatially and temporally overlapping, and decrease with increasing of the spatial and temporal discrepancy (see Chapter 1 for the explanation of the rules). A brief overview about the role of the spatial and temporal rule on the generation of saccades will be here provided.

5.1.4.1 Spatial characteristics of the multisensory integration in saccade generation

Spatial characteristics of the MI in saccade generation were studied both in animals and humans. For example, near-threshold AV stimuli are localized more accurately than V only stimuli by cats (Stein, Hunneycut & Meredith, 1988). The improvement of performance was evident with spatially coincident AV stimuli, and disappeared when A and V were spatially disparate (60° interstimulus distance). In humans the SRT decreases by 20 ms when A and V stimuli are spatially coincident and no reduction of SRT was evident when they were presented in opposite hemifields (Lee et al., 1991).

Hughes et al., (1998) compared the influence of an A stimulus on visually-guided SRT to determine whether the facilitation was due to statistical facilitation or neuronal co-activation (Miller, 1982). With spatially coincident stimuli, the SRT decreased by 50 ms in relation to the baseline and the amount of facilitation exceeded the Race model. When the two stimuli were spatially disparate (with 30° of interstimulus distance), the facilitation was statistical rather that co-activational (i.e., it did not violate the Race model prediction).

Frens, Opstal and Van der Willigen (1995) replicated the 50 ms reduction in SRT with a spatially coincident auditory accessory stimulus and visual target. However, when the auditory non-target stimulus was presented with a spatial separation of 56° from the visual target, the effect disappeared. In their next experiment 2, the authors introduced a

large set of interstimulus distances which allowed them to show a significant correlation between the A-V interstimulus distance and the decrease of SRT: the SRT decreased with about 0.5 ms per degree of stimulus separation. Moreover, no reduction in SRT was evident when the auditory non-target happened to be non-localizable (e.g. low intensity of the relative to background), supporting the idea that the spatial alignment of the stimuli is crucial for the MI (Corneil, Van Wanrooijet al., 2002).

In a more recent study, Colonius and Arndt (2001) studied the SRT to visual stimuli in four visual positions as a function of the position of the auditory stimulus. A MRE for SRT was found in all the MS conditions in comparison to the US conditions. However, the amount of MRE was dependent on the interstimulus distance. For large interstimulus distances (auditory and visual stimuli in different hemifields and an interstimulus distance of 45° or 55° of visual angle) the MRE was of 4.8– 41.2 ms, while the MRE for close stimuli (5° distance) and for coincident stimuli varied between 19 and 63.3 ms.

It is interesting to point out here how there were no evidence of a multisensory depression of the responses even at large interstimulus distances (contrary to what is shown in neurophysiological studies). This point will be reconsidered in the general discussion, after the exposition of my results.

5.1.4.2 Temporal characteristics of the multisensory integration in saccade generation

A few behavioural studies manipulated the asynchrony between auditory and visual stimuli to address the temporal window over which stimuli may interact (Colonius & Arndt, 2001; Corneil & Munoz 1996, Engelken & Stevens 1989, Frens et al., 1995, Ross & Ross, 1981). For the saccadic system, the temporal window is about +/- 100 ms, presumably allowing AV-integration despite differences in retinal versus cochlear transduction times (~ 50 ms and 2-10 ms respectively; Kraus & McGee, 1992).

Diederich and Colonius (2007) measured SRT in a focused attention task with a visual target stimulus and auditory and tactile accessory stimuli at five different SOAs (250, 100 and 50 ms before target onset, simultaneously or 50 ms after target onset). The authors reported that the SRT decreased when the number of non-targets was increased

and when target and non-targets were presented in the same hemifield. Moreover, this facilitation increased when the non-targets preceded the target at different SOA, and then decreased when the non-targets were presented after the target onset.

Colonius and Arndt (2001) have proposed a two-stage model that explains audio-visual interactions in saccades (Figure5.3). In their studies, a visual target was accompanied by an auditory non-target, and the task was performed in a focused attention condition (in which the attention was oriented toward the visual stimulus). Their model, named the TWIN model (Time-window-of-Integration) was firstly developed for audio-visual interactions, and later extended to somatosensory non-target conditions (Colonius & Diederich, 2004; Diederich & Colonius, 2007a). The model allows for both explanations of the multisensory speed-up of the response: statistical facilitation and MI.

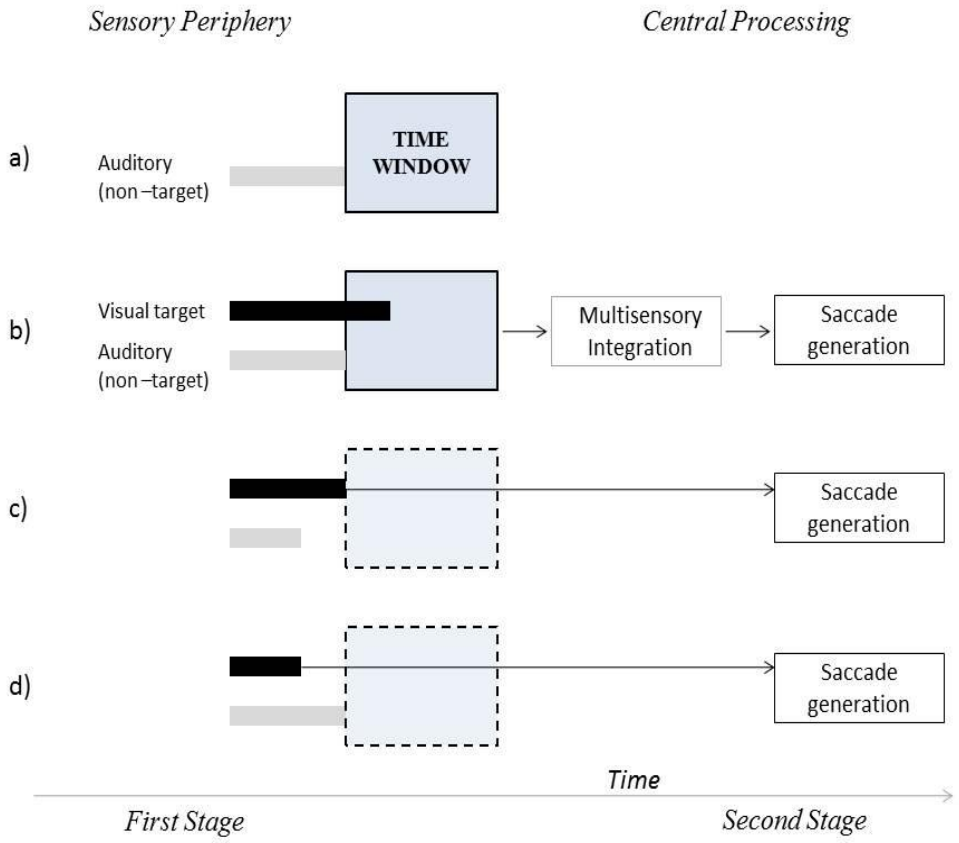


Figure 5.3 *From Colonius and Arndt (2001)*: **A two stage schema of the TWIN model**: a) the auditory non-target opens the time window within which the stimuli will be integrated; b) the visual target falls into the time window and it gets integrated with the auditory stimulus; c) the visual target is the winner of the race but cannot open the MI window; d) the auditory non-target opens the window of integration, but the visual target arrives too late and doesn't get integrated with the auditory stimulus.

In the TWIN model, 2 stages of the AV interaction are assumed:

First stage: This stage refers to very early sensory processing, in which two (or more) signals are processed separately (visual, auditory or somatosensory). Random processing times for visual, auditory and somatosensory stimuli are assumed to be statistically independent. This stage can terminate in two different ways (Figure 5.3): either when the first of the two signals reaches a final centre (Figure 5.3 a); or when the visual stimulus (in a focused attention task) reaches the final stage (Figure 5.3 b).

Second stage: In this stage the stimuli are integrated and oculomotor response is planned. MI occurs only if the accessory stimulus (e.g. audio or somatosensory) is the winner if the "race" in the first stage, when the two stimuli are processed separately (Figure 5.3 a and *b*) and the target (e.g. visual stimulus) arrives within a certain time-window. The time window filters sensory afferents to determine whether they are close enough in time to be integrated. The MI that occurs in this stage can be expressed both as decreasing (enhancement) or increasing (inhibition) of the SRT, depending of the spatial configuration of the stimuli.

This model considers only temporal factors, without taking into account the spatial coincidence. However, it assumes that all the stimulus characteristics that influence their detectability, such as auditory and visual intensity or eccentricity of the visual stimulus, have a direct impact on the first stage of processing, indirectly influencing also influence the second stage.

5.2 The remote distractor effect and Saccadic Inhibition

Aside from the facilitatory effects of MI, the SC has been described as the main structure involved in two inhibitory effects in the saccadic literature: the remote distractor effect (RDE, Walker et al., 1995; Walker, Deubel, Schneider, & Findlay, 1997; Levy-Schoen, 1969; Findlay & Walker, 1999) and saccadic inhibition (SI, Reingold & Stampe, 1999; 2000;

2002; 2004). Contrary to the MI effects reported in previous paragraphs, the phenomena described here refer to the response cost associated with a concurrent stimulus on the SRT toward a target. Moreover, recently the RDE and SI phenomena have been shown to be one and the same phenomenon, albeit typically measured in different ways (Buonocore & McIntosh, 2008). Both effects have been mainly investigated in the domain of the visual perception (with visual targets and visual "distractors"). However, a few attempts to explore a possible multisensory nature of distraction effects have appeared in recent years, and will be shortly described in the next paragraphs.

5.2.1 The Remote distractor Effect

The RDE refers to the phenomenon in which the onset of a simultaneous (or near-simultaneous) distractor at a non-target position increases significantly the SRT to a visual target (Levy-Schoen, 1969; Findlay & Walker, 1999). Levy-Schoen (1969) first showed that SRT are increased (40 ms) when a distractor appear simultaneously in the symmetric contralateral position to the target. In contrast, if the target and the distractor are presented in the same hemifield, no increase in SRT is found but only an influence over the saccade amplitude.

Walker et al. (1995) studied the influence of different SOAs on the RDE. In their experiment, participants were instructed to move their eyes toward a visual target while a visual distractor was presented contralaterally at various SOAs (20, 40, 80, 160 and 240 ms before the target onset, simultaneous with the target or 20, 40, 80, 160 and 240 ms after the target onset). The average SRT was significantly reduced (by 16 ms) when the distractor preceded the target onset (by 160 or 240 ms), indicating a warning effect.

On the contrary, SRTs were prolonged (by 18 ms) when the distractor appeared simultaneously with the target, and the inhibiting effect decreased if the distractor appeared at short intervals (20-80 ms) before or after the saccade target onset. These results were in line with the results of Ross and Ross (1981). Both studies supported the conclusion that the RDE is maximal when the distractor is presented simultaneously with the target or within 100 ms after the target onset.

In a later study, Walker et al. (1997) manipulated the spatial disparity between the visual distractor and a visual target, presenting the distractor at various position starting from the left to the right of the target (up to 10° both on the left or right of fix cross), or in the same spatial position with the target. The magnitude of the RDE decreased monotonically as the distractor was situated more peripherally, whereas it was maximal when the distractor was present at fixation. Moreover, a reciprocal relationship between the effects observed on latencies and on amplitude was revealed: distractors in distant positions affected the SRT but not the amplitude, whereas distractors on nearest position (inside 20°) to the target affected the amplitude but not the SRT.

To sum up, the inhibitory influence of a visual distractor is observed when the distractor is presented contralateraly or at fixation, with the onset of the target preceding the onset of the distractor.

5.2.1.1 Remote distractor effect with auditory distractors

Most researchers have focused attention on the warning effect of an auditory non-target stimulus, expressed as the speeding of SRT when presented with a visual target. The first attempt to describe a potential distractor effect produced by an auditory distractor was by Ross and Ross (1981, 1980). The authors explored the role of the auditory or visual stimuli presented after the onset or offset of the visual target. Two main results were highlighted in their study: the first concern the generality of the warning effect of any accessory stimulus that precede the target (both visual and auditory). The second (more important for the purpose of this paragraph) is the apparent selectivity of the distracting effect, evident only with visual distractors.

In a later study, Corneil and Munoz (1996) compared the RDE across various conditions in which visual and auditory stimuli acted either as saccadic target or distractor. In their experiments, the target and irrelevant cue (either visual or auditory) were presented in the same spatial position (enhancer condition) or on opposite sides (distractor condition), with the SOA varying. All possible combinations of auditory and visual stimuli in target or distractor condition were presented to participants. The authors

found that the greatest inhibition was obtained with visual distractors within 100 ms following an auditory or visual target. A significant, albeit less strong, inhibition was observed when the auditory distractor was presented 100 ms after the onset of a visual or auditory target.

As is evident from the brevity of this review, there are only few studies that have hypothesised a distracting effect of an auditory accessory stimulus. However, the results indicate a lack of inhibition from an auditory distractor or, even when present, such distracting effect is quantitatively different from the effect of a visual distractor.

5.2.2 The Saccadic inhibition paradigm

Saccadic inhibition is a characteristic dip in the saccade frequency distribution (i.e. the frequency of SRTs in each experimental condition) after a transient visual change is presented during the reading or searching task.The paradigm that revealed this effect was described by Reingold and Stampe (1999). In their experiment, during both a reading and scene viewing task, the scene was replaced for 33 ms by a grey field of matching luminance (for the scene viewing task), whereas the text was repleaced by a black flickering square covering the entire screen (in the case of the reading task). The black square produced a dip in the frequency distribution of saccades, starting 60-70 ms after the onset of the flicker and reaching its maximum 90 ms after the onset. This effect was replicated in a more standard saccadic task by Reingold and Stampe (2002), in which participant were required to saccade toward a lateralized target, while a 33 ms white flash was displayed on the top and on the bottom of the screen.

The authors claimed that the timing of the flash relative to the target onset is crucial in order to produce the saccadic inhibition. For example, a flash delay of 50 ms (flash presented 50 ms after the target onset) is expected to produce the maximal inhibition after 100 from flash onset (i.e., 150 ms from target onset). As result, a 50 ms flash delay will produce a dip in the early part of the SRT distribution (short-latency saccade) but not in the later part of the distribution (long-latency saccade). In contrast, a 150ms delay will produce a delay in the later part of the distribution (the latency of the maximal dip will be approximately 250 ms). They assumed that, in order to produce the

maximal saccadic inhibition, the flash delay should be equal to the median SRT minus 100 ms, so that the peak of inhibition impacts at around the centre of the baseline SRT distribution.

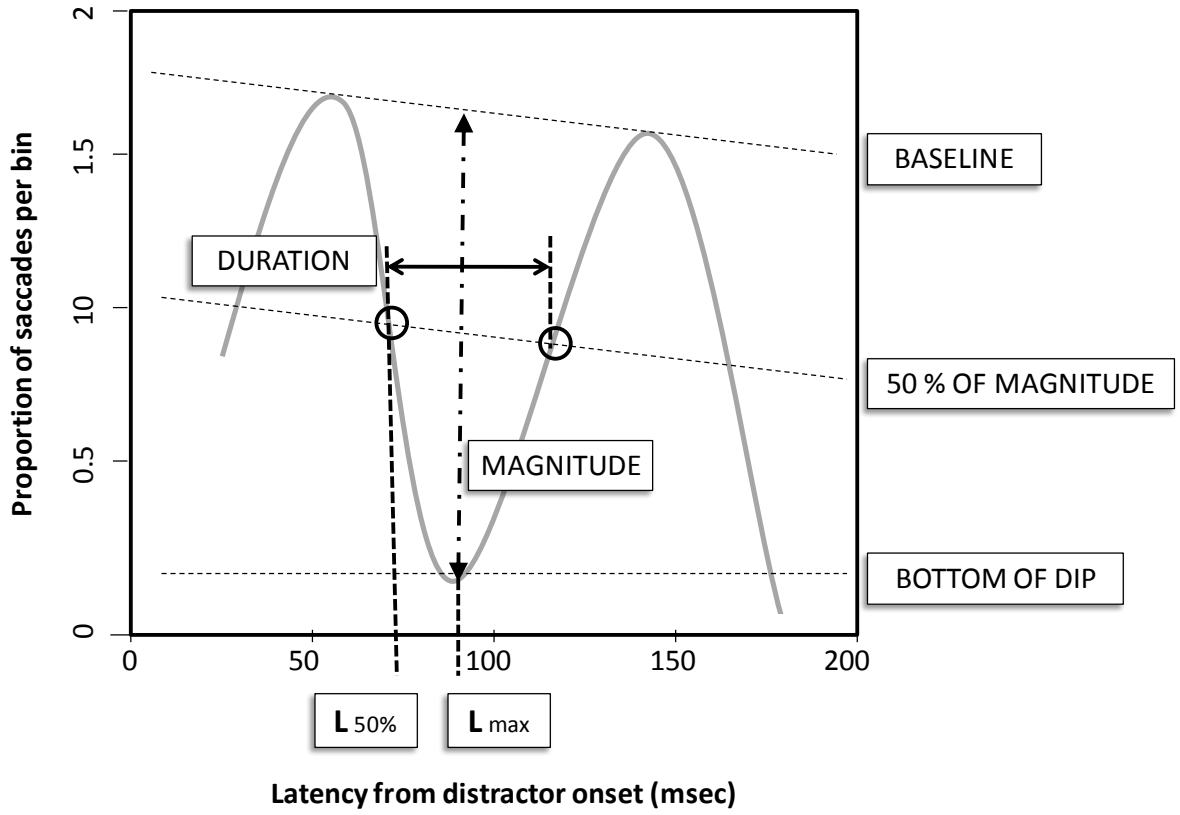


Figure 5.4 *From Reingold and Stampe (2002).* **Schematic representation of quantitative saccadic inhibition measures**: Duration of the dip, Magnitude of the dip, L_{max} - Latency to maximum saccade inhibition, $L_{50\%}$ - Latency to 50% of maximum inhibition.

Reingold and Stampe (2002, 2004) proposed several quantitative measures that characterise the strength and latency of the saccadic inhibition effect (Figure 5.4):

Magnitude of inhibition. The magnitude of saccadic inhibition is defined as the proportion of saccades inhibited during the period of lowest saccadic frequency (i.e., at the centre of the dip), and indicates the strength of the inhibition caused by a certain distractor.

Latency to maximum saccade inhibition (L_{max}). The L_{max} is defined as the time interval from the onset of the distractor to the centre of the bottom of the dip.

Latency to 50% of maximum inhibition ($L_{50\%}$). The L50% is defined as the latency from the onset of the distractor at which inhibition first reach 50% of its maximum strength.

Duration of inhibition. The duration of inhibition is defined as the duration of the period in which inhibition remains above 50% of its greatest strength. The duration is obtained from the difference between $L_{50\%}$ and the end of the period of inhibition.

5.2.2.1 Saccadic inhibition with auditory distractors

Similarly to the RDE, only a few attempts to investigate the influence of a competing auditory distractor in the SI paradigm have appeared in recent years, with contrasting results. Moreover, the first one by Pannasch, Dornhoefer, Unema, & Velichkovsky (2001) examined only the influence of the auditory distractor on the duration of the fixation, while the second one by Reingold and Stampe (2004) considered both the fixation duration and the SRT.

In a picture-viewing task by Pannasch et al. (2001), either a visual (1° black dot) or an auditory distractor (1000 Hz tone) was presented, 100 or 300ms after the onset of a picture. The distribution of SRTs showed an underrepresentation of fixations caused by both visual and auditory distractors that appeared within approximately 100 ms from the onset of the visual, and 80 ms from the onset of the auditory distractor. The authors hypothesized that both the prolongation of fixation and the prolongation of SRT (although not investigated by them) may have a common cause, since the effect that they described (20 ms prolongation of fixation) agrees well with the effect size found in saccade latencies (e.g. Walker et al., 1997).

In a later study, Reingold and Stampe (2004) investigated the influence of an auditory distractor on SI in a reading task. The visual distractor was a black transient image displayed for 33 ms, while the auditory distractor was a 2000 Hz square-wave tone, both lasting for 33 ms. Similarly to Pannasch et al. (2001) the fixation was longer with the visual

transient. However contrary to Pannasch et al. (2001) they did not find an effect of the auditory tone. A similar result was obtained with the SRT: a clear saccadic inhibition dip was evident in the condition with visual transient but no effect of the auditory stimulus.

At this point of research, contrasting evidence does not allow us to fully understand the question of the multisensoriality of the SI phenomenon. However, the multimodal explanation of the effect is consistent with an origin for SI in the SC, since the SC is known as one of the principal structures of the MI. Moreover, it would be interesting to investigate the role of multisensory distractors, which have never been taken into consideration in the literature so far.

5.2.3 Remote distractor effect and Saccadic inhibition: same mechanism?

The studies reported in previous paragraphs highlight abasic difference between RDE and SI paradigms: the RDE has usually been studied in relation to a visual (or in some studies auditory) localized distractor, while SI has usually been studied with a large flash (or a non-localizable auditory stimulus). Besides these differences, both phenomena seem to concern the effect of a distracting event on the saccadic generation. In fact, Reingold and Stampe (2002) in their discussion suggested that the two phenomena may be one and the same effect, and that SI might cause the increased SRT observed in the RDE. This hypothesis was tested by Buonocore and McIntosh (2008). The authors tested the "saccadic inhibition hypothesis", which states that the dip in the SI is time locked to the distracting flash and independent of the temporal delay between the target and the distractor. Following this logic, the amount of RDE should be influenced by the delay between the distractor and the planned saccade instead of the target- distractor delay, and dependent on different portions of the baseline distribution being impacted by distractors appearing at different times.

In order to test this prediction, the authors systematically varied the SOA between the target and the distractor. Their results replicate the RDE obtained by Walker et al., (1995). Moreover, the amount of RDE was consistent with the dip profile of the SI, strongly implying that these phenomena are in fact one and the same. According to this conclusion,

the RDE does not depend on temporal asynchrony between target and distractor but rather on the relation between distractor onset and the baseline SRT distributions.

5.2.4 Neurophysiological basis of Saccadic inhibition

Given that SI has been identified as the functional basis of the saccadic retardation measured as the RDE, they must share a single mechanism. This is consistent with the fact that these two phenomena, considered separately, have been suggested to have similar neurophysiological bases (Findlay & Walker, 1999; Reingold & Stampe, 2002).

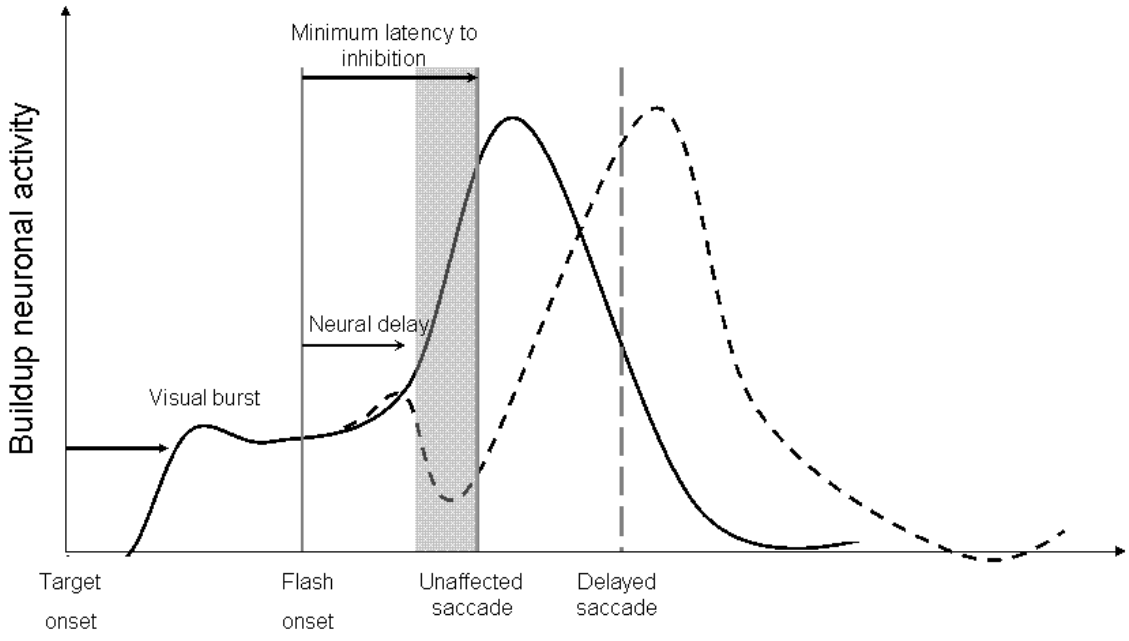


Figure 5.5 *From Reingold and Stampe (2002*). The model shows the hypothetical activity of a build-up neuron in the SC during saccadic inhibition. Two activity patterns are represented: the first corresponds to a saccade unaffected by the distractor (flash) the second corresponds to a saccade delayed due to the inhibition associated with a flash related activity in the SC.

Walker et al. (1995, 1997) observed the largest RDE magnitude when the distractor was at the fixation point and the effect reduced monotonically as the distractor was positioned more eccentrically, concluding that the fixation neurons (which present a sustained activity during active fixation) must play a crucial role in the inhibition effect. The authors suggested that the onset of a target influence the saccade by increasing the activity in the fixation neuron and slowing down the triggering process, and the strength of such an influence depends on the degree of eccentricity of the distractor. Walker et al. (1997) hypothesised that the RDE was the result of competition between the rostral and caudal collicular neurons that are responsible for maintaining fixation and moving the eyes respectively (Munoz & Wurtz, 1993). However, the finding that the latency is increased even at larger eccentricities of the distractor showed that inhibitory effect operates over large areas of the visual field, and are not restricted to the central fovea region, which made the authors advance the idea of an "extended fixation zone".

Another, not necessarily mutually exclusive, explanation claims that the origin of the effect is in the competitive interactions between different movement programs elicited by the target and the distractor (Munoz & Istvan, 1998; Olivier, Dorris, & Munoz, 1999; Rafal, Smith, Krantz, Cohen, & Brennan, 1990; Walker et al., 1997; Walker, et al., 2000). This competition has been hypothesized to operate in the following way: both target and distractor onset may stimulate saccade build-up neurons coding for the location. In that way, they are both competing for target selection, and this competition would generate the delay in triggering the saccade. Olivier et al. (1999) suggested that the source for the RDE is lateral inhibition between subpopulations of build-up neurons in the intermediate layers of the superior colliculus, coding for spatially incompatible saccades. Moreover, this hypothesis is well in accordance to the *saccadic inhibition hypothesis* (e.g. Reingold & Stampe, 2002, 2003).

SImilarily to Olivier et al. (1999), Reingold and Stampe (2002), in the attempt of explaining the SI, proposed two mechanisms of the neural activity associated with the flash transient, both hypothesised to act through inhibitory connections within the intermediate SC. The first mechanism concerns the reduction of pre-saccadic activity in the build-up neurons which response fields corresponds to the required saccadic direction and amplitude, and is the consequence of such inhibitory interactions (Oliver, Doris & Munoz,

1999). The second mechanism involves the stimulation of the fixation neurons by the visual activity associated with the flash, which might inhibit pre-saccadic activity in build-up neurons in the SC (Munoz & Wurtz, 1993 a, b).

Figure 5.5 (Reingold & Stampe, 2002) illustrates the time-course of the second mechanism here described. At the beginning, both patterns show the same activity with a visual burst caused by the build-up cells after the target onset (Munoz &Wurtz, 1995a). After the flash onset, the pattern of build-up neuron activity for the unaffected saccade and the delayed saccade start to diverge, creating the inhibition. The inhibition is maximal if the flash activity starts to influence the build-up neuron prior to the point in which the saccade related motor burst is unstoppable (around 90 ms prior to the execution).

Similarly to Reingold and Stampe (2002) and Olivier et al. (1999), Godijn and Theeuwes (2002) proposed a *competitive integration model*, which assumes that saccade programming is the result of competition between activation at locations represented a common saccade map with a retinotopic representation, in which information coming from different sources is integrated. This model postulates that, when the activation happens simultaneously in two distant locations on the map a mutual inhibition occurs, resulting in the RDE. This occurs because target- and distractor-related activities start a race to reach threshold. The two activated sites inhibit each other so that it takes longer for one of the two to reach the threshold for saccade initiation. On the contrary, if the two activations are spatially close, they combine and result in peak activation somewhere between the two locations. The execution of a saccade is triggered when the activation at a specific location in the saccade map reaches a threshold (see also Trappenberg, Dorris, Munoz, & Klein 2001).

5.2.5 Influence of target and distractor characteristics on Saccadic inhibition

The strength of inhibition (measured either as SI, or as RDE) has been demonstrated by many studies to be influenced by the specific characteristics of both the target and of the distractor. According to the *competitive integration model* (Godijn & Theeuwes, 2002) stimulus properties are thought to influence the speed at which the corresponding activity rises in the saccade map, the peak activity that is reached or the latency of the neuronal response (i.e. the time between stimulus onset and the corresponding neuronal activity in the saccade map). Therefore, characteristics that make a target "highly efficient" (e.g. high contrast, dimension etc...) are thought to enhance one of these aspects of the neuronal activity, and make them more resilient toward a distractor. However, the same logic is applied in the case of the distractor: the more enhanced is the neuronal activity corresponding to some property of the distractor, the more efficient it will be in inhibiting the response toward a target.

For example, it is known that increasing target diameter speeds SRT (Perron & Hallett, 1995). In order to test whether comparable effects are observed also with distractors, Vitu, Lancelin, Jean, and Farioli (2006) presented central distractors of various lengths. The authors reported that the RDE increased with increasing distractor length for central distractors (strings of letters) that extended between 0.31° and 2.79°, confirming that large stimuli can act both as efficient targets and efficient distractors.

In the study by Ludwig, Gilchrist and McSorley, (2005), the spatial frequency of target and distractor were varied simultaneously, using Gabor patches. Low spatial frequencies were shown to enhance neuronal response more than high spatial frequencies (see also Ludwig, Gilchrist, & McSorley, 2004). Interestingly, instead of the expected main effects of target and distractor spatial frequency, they found an interaction: low to medium (up to 4 cycles/deg) spatial frequency distractors inhibited the saccade toward the target equally, regardless of the target spatial frequency. The inhibition caused by high frequency distractors was weaker in comparison to low frequency distractors, except when the target spatial frequency was high as well.

On the basis of their results, the authors hypothesised an expanded model of the target-distractor competition. In fact, the data from their experiments showed that the basic assumption, of a weak distractor (i.e. high spatial frequency patch) inhibiting less than a strong distractor (i.e. low spatial frequency patch), does not always hold. They suggested that the efficiency of the target or distractor cannot be the only factor influencing the RDE, but possible interactions before the final level of saccade triggering may occur. For example, the activity generated by the target and distractor may result in a channel overlap (e.g. a channel specific for spatial frequencies). Therefore, a two stage mechanism for the RDE is suggested: In the first stage the responses to a target and distractor are reduced depending on the extent to which they activate the same channel. In the second stage this channel specificity is lost. The stimulus activations compete with each other at this point, and the strength of this competition depends on the magnitude of the neuronal responses only, regardless of which channel mediated these responses.

Apart from size and spatial frequency, the contrast of the target and distractor was also investigated in a study by Born and Kerzel (2008). Since saccades are known to be faster to high-contrast than to low-contrast targets (Ludwig et al., 2004; White, Kerzel, & Gegenfurtner, 2006), Born and Kerzel (2008) manipulated the contrast of both target and distractor simultaneously, with distractors presented either peripherally or in central positon. Contrary to their expectations, the results showed that the strength of RDE was not influenced by distractor contrast, especially when it was presented centrally. In fact, the authors reported the opposite pattern of data: the inhibition was stronger with high-contrast targets, while high-contrast distractors did not produce stronger inhibition in comparison to low-contrast distractors.

Interestingly, they found that the RDE was rather influenced by the SRT elicited by the target, showing that shorter SRT leaded to stronger RDEs. The authors argued that the SRT to a target is crucial in determining whether distractor-related activity will temporally overlap with target-related activity, since the temporal coincidence of the two neuronal activities is essential to inhibit each other. In fact, high contrast of the distractor will result in a faster activation and in the temporal advantage over the activity elicited by that target. In that case, the inhibition will be weaker with high contrast distractors because of the increase in the temporal delay between the distractor and the target activity. Further

explanation for the temporal account will be given in the general conclusion paragraph of the present chapter.

In conclusion, the general rule that seems to emerge from the studies reported in this paragraph seems to be as follows: the characteristics of the target that influence the SRT in absence of a distractor (either by speeding up or slowing down the SRT), seems to influence the resilience of that target toward a distractor. Conversely, the same characteristics influence also the strength of a distractor in inhibiting (or slowing down) a saccade toward a stimulis. This is, however, the case only when the target and the distractor are coincident in time, as suggested by Born and Kerzel (2008). Moreover, the featural similarity between target and distractor seems to influence the strength of the inhibition as well (Ludwig, et al., 2005).

5.3 Audio-visual interaction in saccade generation

5.3.1 Rationale of the experiments

Saccadic inhibition offers an interesting paradigm within which to investigate the interaction between auditory and visual stimuli in the generation of saccades.

Since the SI and the MRE are phenomena that are thought to take part in the same neuronal structure (the SC), the general aim of the series of experiment presented in this chapter is to investigate the relation between these effects and to understand whether and how they influence each other.

Similarly to other studies described in paragraph 5.2.5, properties of target and distractors will be manipulated in Experiments 7, 8, 9 and 10. If we assume that the competition between target- and distractor-related activities modulates the SI, as reported in other studies described in the previous paragraph (Godijn & Theeuwes, 2002; Kopecz, 1995; Trappenberg et al., 2001), we should expect to find stronger resistance with AV targets toward a distractor (due to the stronger neuronal activity), and stronger inhibition

with an AV distractor. In order to compensate for time delays between target and distractor (due to the different time needed for the processing of A, V and AV targets or distractors), and in order to ensure the maximal inhibiting effect (Reingold & Stampe, 2002), the onset of the distractor is time-locked to the median time of saccade initiation, and presented 90 ms after the median SRT of each participant separately. In this way, any difference in the RDE coused by the multisensoriality of the distractor or of the target will be due to factors different from the temporal delay between the target and distractor-related activity. For example, I assume that the AV integration may result in a more robust representation of the target, making it more resistant toward distractors, and the same should be true for the AV distractors.

In Experiment 7, the multisensory nature of the SI effect is tested. In particular, the effect of visual, auditory and audio-visual distractors (contralateral to the target) is compared. In this experiment, I aim to test whether it is possible to observe an inhibition with auditory distractors, and if the inhibition with audio-visual distractors is stronger than the inhibition with the two unisensory distractors (audio and visual).

In Experiment 8, the same research question is studied, using a different arrangement of stimuli. In fact, large flashing zone has been used instead of contralateral distractors, similarly to the stimuli used by Reingold and Stampe (2002). Moreover, the nature of the predicted MI in AV distractors is addressed, using the Race model analysis.

In Experiment 9, three different types of targets are presented with a contralateral visual distractor. The main prediction is that bi-modally (AV) guided saccades will be more resistant to a concurrent visual distractor, due to the enhancement given by the multisensoriality. Moreover, the influence of a visual distractor to the auditory guided saccade is investigated, and compared to the visually and bi-modally guided saccades. The same research questions are then addressed in Experiment 10, with some methodological improvements to the stimuli (large flashing zone used instead of lateralized distractors, similarly to Experiment 8).

5.3.2 Experiment 7

The influence of an audio, visual, or audio-visual distractor on the visually-guided saccades is investigated. In particular, it is predicted that, if the integration among A and V stimuli produce a stronger response, the AV distractor will cause stronger inhibition in comparison to audio or video only distractors.

5.3.2.1 Method

Participants

Eight students (age range 20-25) from the University of Edinburgh participated in the experiment. All had normal hearing and normal or corrected-to-normal visual acuity, gave their informed consent prior to the beginning of the experiment and received course credits for their participation. This experiment was conducted in accordance with the 1964 Declaration of Helsinki, with the approval of the Ethics Committee of the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh.

Apparatus and stimuli

Visual stimuli were displayed on a 19 inch CRT monitor (1024 x 768 pixels) driven by a Pentium IV processor at 120 Hz. A chin rest was used to fix the participant head 80 cm from the monitor. Eye movements were recorded by monitoring the position of the pupil with the EyeLink1000, at a sampling frequency of 1000 Hz. The target of the saccades was a white 0.8° x 0.8° square presented on a black background, at a distance of 9.9° from the white fixation cross displayed at the horizontal centre of the screen. In some trials a distractor identical to the target was presented at the same distance to the left side of the fixation cross. Two loudspeakers were positioned on the left and right top edge of the screen, aligned with the target stimulus at the right and the distractor to the left. In order to reduce the vertical difference between the visual and auditory stimuli, the visual stimuli and fixation cross were presented at the top of the screen: 1.5° from the top of the screen

and 4° from the centre of the loudspeaker. The distractors could be either visual (white square), auditory (1000 Hz pure tone), or audio-visual (white square and auditory tone presented simultaneously). The target and the distractors lasted for 700 ms. The onset of the distractor were calculated according the following formula: ONSET= median SRT – 90 ms (Figure 5.6; Reingold & Stampe, 2002). This formula has been adopted according to the method of Reingold and Stampe (2002), which assumes that the maximum inhibitory effect of the distractor will be evident ~90 ms after the distractor onset. Therefore, the median SRT for every subject was calculated, and 90 ms was subtracted from this value. In this way, the distractor onset was time-locked to the expected time of saccade initiation at a per-participant level.

Procedure

Participants performed a preliminary session (50 trials) with the target only condition in order to compute the median SRT for each participant individually. The median SRT of each individual was used to calculate the distractor onset, according to the formula by Reingold and Stampe (2002). After the preliminary session, the main experiment was performed.

Each trial started with a central 0.8° x 0.8° fixation cross at the top of the screen. After a random interval (500-1000 ms) the target appeared on the right of the fixation cross. In the distractor trials, the visual, auditory or audio-visual distractors appeared to the left of fixation (Figure 5.6). Participants were required to fixate the cross and to move their eyes to the target as soon as it appeared on the right side of the screen, and to ignore the distractors on the left side of the screen. The experimental conditions were thus the following: visual target alone (T), target with a visual (T_VD), audio (T_AD) or audio-visual (T_AVD) distractors.

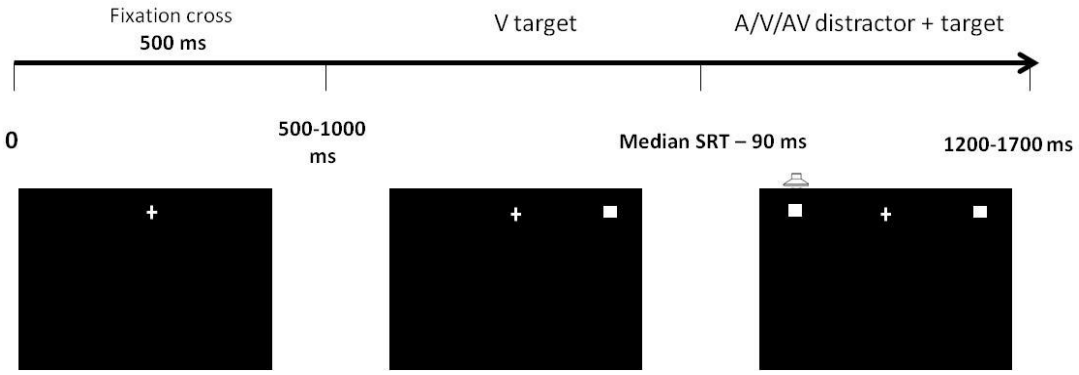


Figure 5.6 Experiment 7, stimuli and timetable of a typical experimental distractor trial (not in scale). Each trial started with a fixation cross, followed by a variable duration ranging from for 500 – 1000 ms, and the presentation of the target (a white square on a black background) always on the right top side of the screen. In the "distractor" trials, one of the possible distractors wee presented (A,V or AV), either on the left top side of the screen or from the left loudspeaker, or both.

Participants completed 800 trials, divided in four sessions, each composed of 10 blocks. One block contained 20 trials of each trial type, with a total of 80 trials per block, presented in a random order. Each trial began with drift correction and a tone accompanying the onset of a 0.50° central fixation cross.

The overall duration of the experiment was 90 min. A break was given to participants at the end of every session, or every time they needed to rest. A three point calibration was run at the beginning of every session and after five consecutive blocks. The experiment was performed in a darken room.

Data screening

Saccades to the left (1.2%), saccades of less than 1° amplitude (3.5%), and saccades with latency above 70 ms (1.5%) or over 500 ms (3.2%) were excluded

5.3.2.2 Results

SRT analysis[2]

For each participant, the *Distractor cost* was calculated subtracting mean SRTof the baseline T only condition from each Distractor condition. In this way were obtained three measures of the "cost" that every distractor type (A, V and AV) causedto the T only baseline. The distractor cost values of each subject were entered into a one-way within subject ANOVA that revealed a significant main effect of Distractor type ($F_{2,14}$= 3.98, p< 0.05). Planned comparisons revealed a difference between the AD and AVD condition ($F_{1,7}$= 6.4, p< 0.04), but no difference between VD and AVD ($F_{1,7}$= 3.0, p< 1), and no difference between A and V ($F_{1,7}$= 1.7, p= 0.2).

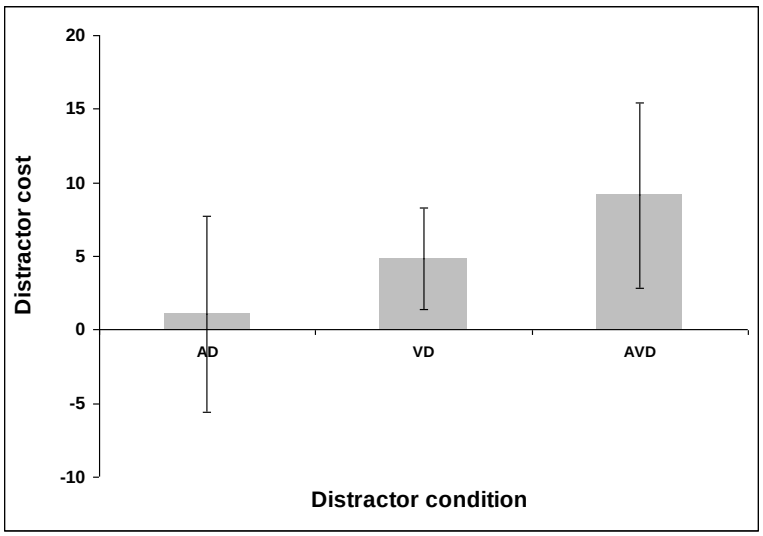


Figure 5.7 Experiment 7, mean distractor cost. Mean Distractor cost and 95 % confidence interval was calculated for each distractor condition: Audio distractor (AD), Visual distractor (VD), and audio-visual distractor (AVD).

[2] All the analysis in the present chapter are performed on RT instead of the 1/RT transformation (adopted in the previous chapters), in order to follow the procedure usually reported in the literature on the RDE. However, I performed a control analysis (not reported here) on 1/RT as well, and found no difference in the main effects or interactions.

A series of t tests against 0 revealed a significant difference only in the AVD condition ($t(7)= 3.35$, $p< 0.01$) whereas no difference were evident for the AD and VD condition ($t(7)= 1.16$, $p= 0.3$ and $t(7)= 1.90$, $p= 0.1$ respectively).

SI analysis

A SI analysis was performed on the data. To provide an illustrative demonstration of the impact of different distractors on the magnitude of the dip, difference histograms were derived. For each participant and each experimental condition, the percentage frequency histogram of SRT was computed with a bin-width of 4 ms. A seven-point-average smoothing function was applied in the analysis. Difference histograms were obtained by subtracting bin-by-bin the histogram of the baseline Target only condition from the three distractor distributions. Difference histograms were aligned to distractor onset and averaged across participants for each distractor condition separately. This averaging was performed just for illustrative purposes, in order to provide a graphic representation of the three dip profiles (Figure 5.8); note that, in some cases, the illustrative representation may differ slightly from the values reported in the analysis.

Table 1 report the means of the principal SI measures, described in the 3.3.2 paragraph.

The *magnitude of the dip* was derived from the minimum of the difference distribution of each distractor condition from the baseline for every participant separately.

The *L50%* is the latency from the flash onset at which inhibition first reaches 50% of its maximum strength.

The *Lmax* (latency to maximum saccadic inhibition) is equivalent to the computed latency of the dip.

The *duration* corresponds to the period during which inhibition remains above 50% of its maximum strength, and is calculated subtracting L50% from its counterpart part on the other side of the dip (L50% > Lmax).

The *Recovery time* is the latency that corresponds to the difference between the latency of Lmax and the latency of the maximum peak of the difference distribution

Table 1. Measures of saccadic inhibition (mean and SD) in Experiment 7 for each of the three distractor types: the magnitude of inhibition, the latency to the maximum inhibition (L_{max}), the duration of the dip and the recovery time.

Distractor type	Magnitude (%)	Lmax (ms)	Duration (ms)	Recovery time (ms)
V	-1.7 (0.8)	86.5 (17.9)	44.9 (26.4)	53.5 (29.6)
A	-1.2 (0.8)	67.0 (30.7)	62.0 (40.8)	41.5 (24.8)
AV	-1.6 (0.6)	77.5 (29.1)	53.3 (43.6)	56.5 (29.6)

The dip magnitude was extracted from difference histogram for every distractor condition. Three t test against 0 confirmed the significance of the dip for every distractor condition: V (-1.7), $t(7)= 6.7$, $p< 0.001$; A (-1.2), $t(7)= 4.6$, $p< 0.002$; AV (-1.6), $t(7)= 8.1$, $p< 0.001$). However, the three dip profile did not differ from one another in the *dip magnitude* ($F_{2,14}$=1.96, $p= 0.2$), *Lmax* parameter ($F_{2,14}= 2.8$, $p= 0.09$), *Duration* ($F_{2,14}= 2.02$, $p= 0.2$) or the *Recovery time* ($F_{2,14}= 0.6$, $p= 0.5$) (Table1).

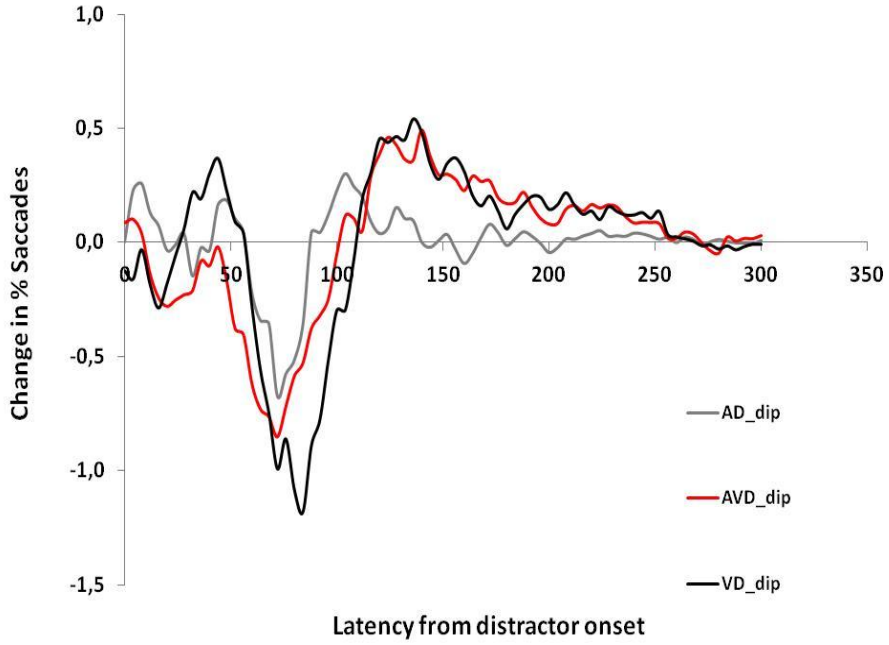


Figure 5.8 Experiment 8, Saccadic inhibition dip. Difference histogram for each of the distractor condition derived from the bin-by-bin subtraction of Target only histogram from distractor histograms for the three distractor conditions: V, A, AV.

5.3.2.3 Conclusions

The results from Experiment 7 have partially answered our initial questions, while other questions still remain open, perhaps due to methodological imperfections that will be discussed shortly.

Despite its limitations, some results are clear from the experiment. First of all, a significant RDE was obtained with the AV distractor, whereas no effect were evident with V or A distractors alone, suggesting that a bimodal distractor can be more effective in inhibiting visually-guided saccades in comparison to the two unisensory components alone. However, I have to point out that the result was intermediate, since the difference with the V distractor alone was not significant, whereas was significant in comparison to the A distractor alone.

The second result, based on the SI analysis, confirmed that either the V, A or AV distractors produced an inhibition, equal for the three types of distractor. The main limitation of this experiment was that, although robust, the absolute magnitude of the three SI dip were small, and therefore there was little scope for significant modulation by distractor modality. Both the eccentricity and the size of the distractor might influence the strength of the Inhibition. In particular, it is known that the probability of SI to occur for a given saccade increases logarithmically with the size of a contralateral distractor (Walker et al., 1997for RDE; Buonocore & McIntosh, 2011 for SI). Therefore, a better choice would be a larger distractor and shorter distance between target and distractor. In the case of the auditory distractor, the 1000 Hz pure tone should be replaced with a white noise since it is well known that our ability to localize a sound decreases with decreases in bandwidth for both binaural and monaural listening conditions (Langendijk & Bronkhorst, 2002; Hofman & Van Opstal, 1998; Langendijk et al., 2001).

In conclusion, the fact that a significant inhibition was obtained with SI analysis, whereas no effect was evident with SRT analysis is an indication that distributional (i.e. dip) analysis is a more sensitive and revealing way to investigate distractor effects than is mean SRT. This was pointed out also by Buonocore and McIntosh (2008), who warned that the investigation of distractor effects should be carried out by an SRT distribution analysis rather than by an analysis of central tendency measures, due to the bimodality induced by distractors.

To sum up, the results of the Experiment 7 are intermediate; albeit weak, a suggestion of an effect is present with SRT, but no clear difference in SI parameters. Therefore, further investigation is needed to clarify the pattern of results.

5.3.3 Experiment 8

The hypothesis from Experiment 7 was re-tested in Experiment 8. However, in order to resolve the methodological imperfections of Experiment 7, some stimuli have been modified. The major problem in Experiment 7 was the relatively weak distraction effect in all the experimental conditions, as reported in both the RDE and SI analysis. In

order to potentiate the efficiency of the distractors in Experiment 8, a large flashing zone (cf. Reingold & Stampe, 2002) was used instead of a small lateralized square, and a generalized white noise burst was used instead of the 1000 Hz lateralized tone.

5.3.3.1 Method

Participants

Fifteen participants who had not taken part in Experiment 7 were tested (seven female and four male, age range 20-26). All had normal hearing and normal or corrected-to-normal visual acuity. They gave their informed consent prior to the beginning of the experiment and received course credits, or were paid for their participation.

Apparatus and stimuli

The same apparatus was used as in Experiment 7. The target of the saccades was a 0.3° x 0.3° black square presented on a grey background (in order to replicate the original setting from Reingold & Stampe, 2002), 7° to the right or left of the fixation cross, displayed at the horizontal centre of the screen. In “visual distractor” trials (flash trials), a white flash region (25.6° x 4.2°) was displayed for 30ms at the bottom and at the top of the screen. In “auditory distractor” trials a white noise general burst (80 dB) was presented for 30ms from two loudspeakers situated at the bottom and at the top of the screen. In “audio/visual distractor” trials, a white flash and a white noise were presented at the same time after the same interval as in flash and auditory trials.

Procedure

The participants performed the preliminary session, similarly to Experiment 7. At the beginning of each trial, participants were required to fixate the cross (0.8° x 0.8°) at the centre of the screen. The fixation cross was presented for 500ms (see Figure 5.9). The

target was then randomly presented to the left or right of the fixation (approximately 7° from the fixation cross), in order to prevent the response from becoming too highly automatic. Participants were required to generate a saccade toward the target as soon as it appeared on each side of the screen, and to ignore the flash/sounds in the distractor conditions.

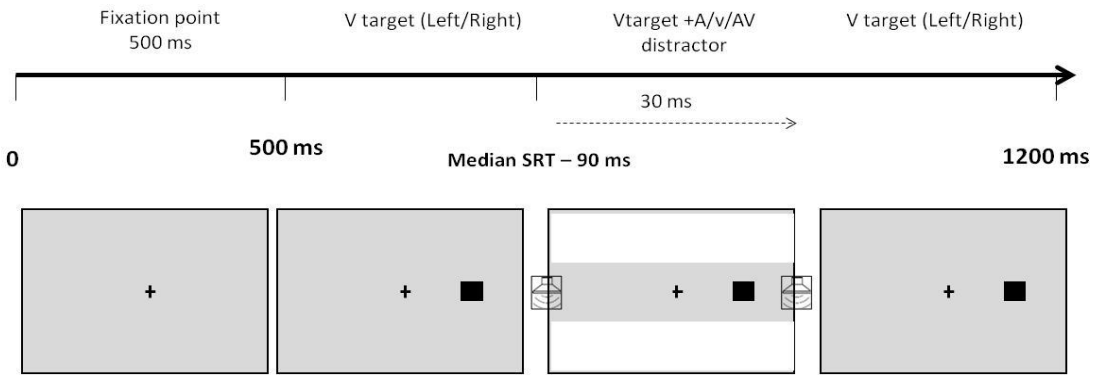


Figure 5.9 Experiment 8, stimuli and timetable of a typical trial (not in scale). Each trial started with a fixation cros lasting 500 ms, followed by the presentation of the target (a black square on a grey background) randomly on the right or left side of the cross, in the central portion of the screen. In the "distractor" trials, one of the possible distractors were presented (A,V or AV). The V distractor was a flashing bright zone on the top and bottom of the screen, lasting for 30 ms. The A distractor was a white noise burst originateing from the two loudspeakers positioned in the central part of the screen, on the left and right side. The AV distractor consisted of the flashing zone and the white noise burst occurring at the same time.

The experimental design consisted of four conditions, presented randomly to the participants: a target only condition (T), in which the target was presented randomly on the left or right of the fixation cross; a visual distractor condition (VD), with a bright flashing visual transient abruptly presented on the top and on the bottom of the screen; an auditory distractor condition (AD), with a white noise burst presented simultaneously from the two loudspeakers on each side of the screen and an audio-visual distractor condition (AVD), in which both the visual flash and the audio burst were presented simultaneously.

To provide sufficient numbers of trials for a SI analysis of SRT distribution, 200 trials per condition were presented (total number of trials= 800), divided into four experiment

blocks, each composed by 200 trials, 20 of each trial type. The overall duration of the experiment was 60 minutes.

Data screening

Saccades directed downward or upward (2.68%) or having less than 1° of amplitude (1.97%), and saccades with latency above 70 ms (1.75%) or over 500 ms (2.07%) were excluded. Participants with more than 15% misses were excluded from the analysis (subject number 2, 8, 10 and 15).

5.3.3.2 Results

A Side (left, right) x Condition (T, AD, VD, AVD) ANOVA performed on SRT revealed a significant effect of Condition ($F_{3,30}$= 17.6, p< 0.001), a significant effect of the Side of presentation ($F_{1,10}$= 3.20, p= 0.05) with saccade oriented to the right being faster than those oriented to the left; and no effect of the Side x Condition interaction ($F_{3,30}$< 1).

SRT analysis

The Distractor cost was calculated for each participant separately, similarly to the analysis in Experiment 7. As evident from Figure 5.10, the overall magnitude of the effect in the three distractor conditions was higher than in Experiment 7, confirming that the more salient distractor produced large effects, and providing more potential scope for identifying differences between conditions.

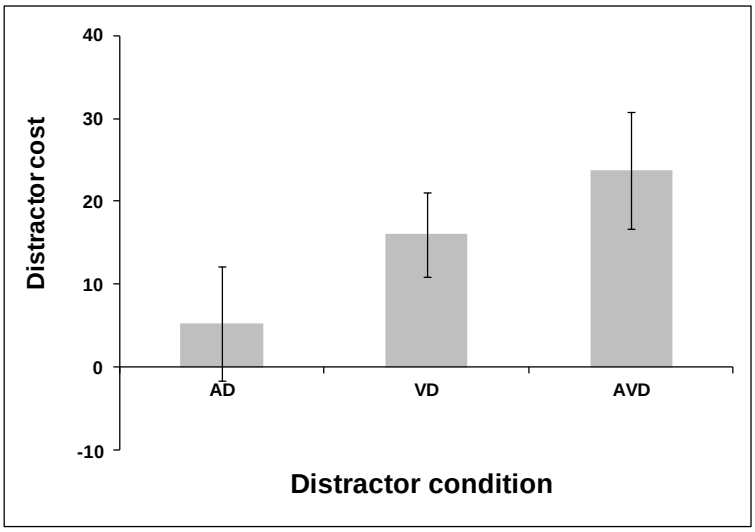


Figure 5.10. **Experiment 8, Mean Distractor cost** and 95% confidence interval were calculated Visual distractor (VD), audio distractor (AD) and audio-visual distractor (AVD) condition.

A main effect of Distractor condition was revealed by a one-way ANOVA ($F_{2,20}$= 9.42, p< 0.002). Given the strong predictions I had about the superiority of the AVD condition over AD and VD, a planned comparison analysis was carried out showing that the AVD distractor cost was stronger than both the VD ($F_{1,10}$= 4.9, p< 0.05) and AD ($F_{1,10}$= 16.7, p< 0.002) distractor conditions. All the three distractor costs revealed a significant difference against 0 (t(10)= 2.5, p< 0.05; t(10)= 5.1, p< 0.001 and t(10)= 7.2, p< 0.001, for the VD, AD and AVD distractor conditions, respectively.

SI analysis

The same procedure as in Experiment 7 was adopted for the SI analysis. For illustrative purposes only, Figure 5.11 represents the three dip profiles (VD, AD, AVD) averaged across participants.

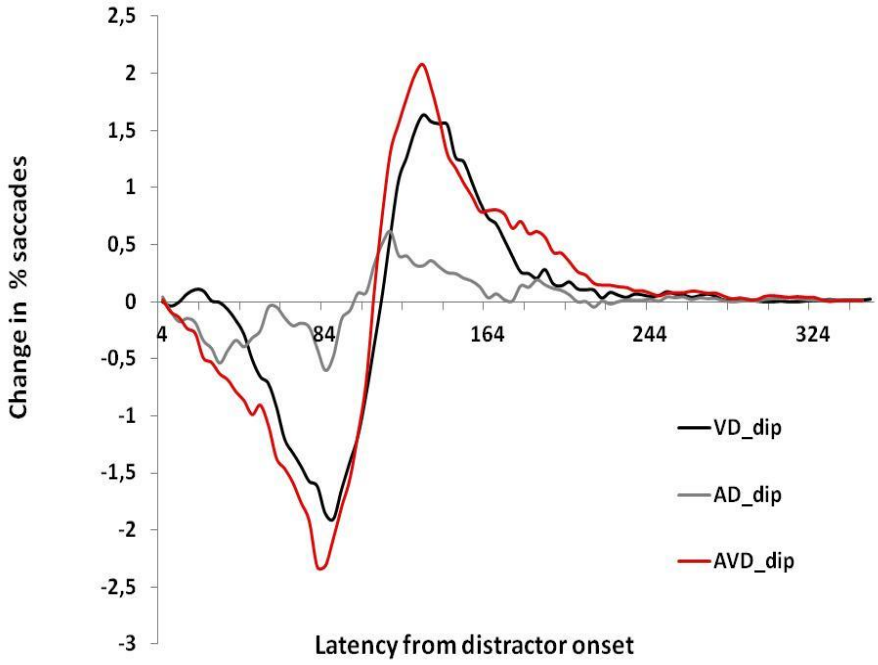


Figure 5.11 Experiment 8, Saccadic inhibition dip. Difference histogram for each of the distractor condition derived from the bin-per-bin subtraction of Target only histogram from distractor histograms for the three conditions.

Six dip parameters were extracted for every dip condition in every subject (Table 1).

Table.2. Measures of saccadic inhibition in Experiment 7 for the three distractor types: the magnitude of inhibition, the latency to the maximum inhibition (Lmax), the duration of the dip (composed by the latency to 50% of the Lmax (L50% <Lmax), and 50% of the latency after the Lmax (L50% >Lmax)), and the recovery time.

Distractor type	Magnitude (%)	Lmax (ms)	L50% (>Lmax)	Recovery time (ms)
V	-3.5 (1.9)	86.5 (30.9)	17.1 (5.2)	53.3 (19.1)
A	-1.9 (0.3)	63.3 (24.5)	12.1 (4.1)	52.3 (20.5)
AV	-3.8 (2.0)	81.8 (38..3)	17.2 (9.1)	58.7 (32.4)

All the parameters, extracted for every single participant, were entered into repeated-measures ANOVA per condition.

There was a significant effect of distractor condition on the Magnitude of the dip: $F_{2,20}$= 8.42, p< 0.002.

The post-hoc analysis revealed significant difference among all the distractor conditions: V vs. A, $t(10)= 2.75$, $p< 0.02$; V vs. AV, $t(10)= 2.53$, $p< 0.03$; A vs. AV, $t(10)= 3.04$, $p< 0.01$. A series of t test against 0 were calculated in order to test the significance of the dip for each distractor condition: V (-3.5), $t(10)= 5.68$, $p< 0.001$; A (-1.88), $t(10)= 22.63$, $p< 0.001$; AV (-3.78), $t(10)= 6.76$, $p< 0.001$).

A significant main effect of distractor condition was obtained for *duration* ($F_{2,20}= 18.4$, $p< 0.0001$), confirming that dip duration was longer in the V and AV conditions thanin the A condition (pairwise comparisons, V vs. A $t(10)= 4.17$, $p< 0.002$; AV vs. A $t(10)= 5.78$, $p< 0.001$) whereas no significant difference were obtained between A and AV condition ($t(10)= 1.91$, $p= 0.09$).

A significant effect of distractor condition was also found for *Lmax* (A $F_{2,20}= 9.49$, $p< 0.01$), but was mainly determined by the difference between A vs. AV (pairwise comparisons, $t(10)= 4.08$, $p< 0.05$) and A vs. V ($t(10)= 3.11$, $p< 0.05$), while no difference were evident between V and AV ($t(10)< 1$), indicating that overall the latency to the maximal saccadic inhibition were the same in V and AV conditions.

Race model analysis

The result from the SRT and SI analysis supported my hypothesis of an enhanced inhibition with multisensory distractors showing both a greater retardation of SRT and an increase of the dip in the saccadic inhibition. Together, these results show that both the auditory and visual distractors contributed to the multisensory inhibition. A further step in the analysis is the definition of the process of integration between the auditory and visual distractor: do they interact, or they both act as individual distractors without being integrated at any stage of the generation of the response?

This question is generally answered by the Race model analysis that tests whether reduced reaction time (or enhanced inhibition) when more than one sensory modality is used is the result of MI (i.e. does it exceed the statistical advantage expected from having multiple sources of information). If the multisensory effect exceeds the upper limit of the advantage (or disadvantage) predicted by the race model, we can conclude that the two

signals interact in the generation of the response rather than to act independently. Since the Race model is generally applied to test the level enhancement of the response (see also Chapter 4) rather than inhibition, an alternative version was derived for the purpose of our study:

$$\text{Strength of inhibition AVD} = A_{\text{inhibition}} + V_{\text{inhibition}}$$

For a detailed explanation of the procedure adopted see APPENDIX of the present thesis.

As shown in Figure 5.12 the AVD condition was slower than AD and VD conditions.

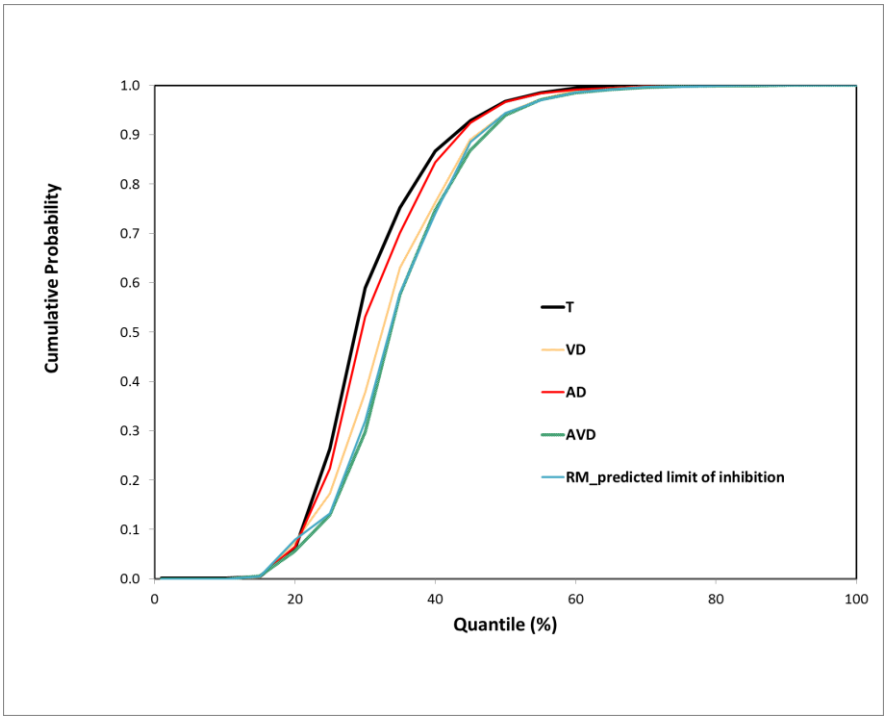


Figure 5.12 Experiment 8, Empirical cumulative distributions for the group of 11 participants in the four experimental conditions and cumulative distribution of the limit of inhibition predicted by the Race model.

However, it is clear from Figure 5.12 that there was no difference between the empirical AV cumulative distribution and the RM predicted limit of inhibition, and no

further analysis was performed on the data. The absence of violation of the RM of the inhibition allowed me to conclude that the A and V components of the AV distractor did not co-activate the inhibiting response.

5.3.3.3 Conclusions

In Experiment 8, the major methodological imperfections from Experiment 7 were remedied by changing the strength of the distractor, as evident from the overall larger distractor cost in all three distractor conditions. In fact, a consistent result in literature regards that modulation of the SI effect by both the dimensions of the distractor (Sumner & Bompas, 2009; Buonocore & McIntosh, 2008; Edelman & Xu, 2009) and its saliency, with bigger and more salient distractors producing stronger SI. The same logic was applied to the auditory distractor, assuming that a white noise would represent a stronger interfering event comparing to a pure tone.

The analysis on SRT confirmed significant inhibition with V, A and AV distractors. The RDE was stronger in the case of the AV distractor, demonstrating that the multisensory enhancement is expressed not only in the strengthening of a target-directed response but also in the strengthening of the inhibition of a response.

The second main result concerns the SI paradigm. A stronger dip was obtained with the AV distractor in comparison to the V and A distractor. This result can be explained by the fact that, in general, a more salient distractor produces a stronger inhibition, as already demonstrated in the case of visual distractors (Born & Kerzel, 2008). Finally, the Race model analysis was applied to the data in order to determine the nature of the AV interaction in the inhibition of the saccade. The AV distractor profile did not violate the race model, suggesting that the two stimuli did not interact at the neuronal level, and that the stronger inhibition was due merely to statistical facilitation (of inhibition in this case).

5.3.4 Experiment 9

As a general rule, we can say that MI enhances a response of an observer, either at a neuronal level by increasing the firing rate or at behavioural level by improving the latency and/or accuracy of detection. In Experiment 9 the question I wanted to answer is whether this enhancement can boost the capacity of saccadic preparation to resist SI. For this reason, three different targets were presented to participants (audio, visual and audio-visual), and the influence of a visual distractor was compared across the three target conditions. According to the SIH, however, the principal factor affecting the strength of inhibition is the delay between the distractor onset and SRT distribution. Given the fact that the latency distributions will differ for the three target types (A,V and AV), presumably being faster for AV targets in comparison to A and V targets, different distractor onset will be calculated for each target SRT distribution. Time-locking the distractor to each distribution, any difference in the dip profile will be attributed to qualitative differences between the three types of saccades (visually, auditory or audio-visually triggered saccades), and not to differences in distractor timing relative to the baseline (target only) SRT distribution.

5.3.4.1 Method

Participants

Eleven participants were tested (six female and three male, age range 20-26). All had normal hearing and normal or corrected-to-normal visual acuity (by self-report), gave their informed consent prior to the beginning of the experiment and received course credits, or were paid for their participation.

Apparatus and stimuli

A fixation cross 0.5° x 0.5° was presented at the centre of the upper part of the screen (2° from the top of the screen), with two empty squares (1.65° x 1.65 °) placed to the right and left of the fixation cross. The empty squares indicated to the participant the possible locations of the visual target and served as place-holders in audio target conditions. The distance between the target and the fixation cross was 5.5° of visual angle. The visual target was a black square (1.5° x 1.5° of visual angle), presented at the centre of the empty square on the right side of the fixation cross (Figure 5.13). The auditory target was a white noise burst (150 ms, 80 dB) presented from a loudspeaker situated above the position of the visual target (Figure 5.13). The audio-visual target was determined by the simultaneous appearance of the visual and auditory targets. The distractor was always the same in all the target conditions: a black square identical to the visual target, presented at a distance of 5.5° of visual angle on the left side of the fixation cross. The onset of the visual distractor was determined individually per participant for each of the three target conditions, and was calculated by subtracting 90 ms from the average median SRT for that target condition (Reingold & Stampe, 2002). The target and distractor disappeared from the screen 700 ms after the target onset.

Procedure

At the beginning of each trial, participants were required to fixate the fixation cross, followed by the random presentation of one trial type. The experimental design consisted of six experimental conditions: Visual target only condition (V), in which only the visual target was presented on the right of the fixation cross; auditory target only condition (A), with the auditory target presented from the right loudspeaker; Audio-visual target only condition (AV), with the visual and auditory target presented simultaneously on the right of the fixation cross; and the three distractor conditions: visual target with visual distractor condition (V_VD), auditory target with visual distractor condition (A_VD) and audio-visual target with visual distractor condition (AV_VD), in which the distractor was presented on the left side of the fixation cross.

Since the target and the distractor were similar, the participant were instructed to expect the target always on the right side of the fixation cross, since the spatial indication was the only parameter that allowed them to discriminate the target from the distractor. However, given the absence of spatial uncertainty of the target, a temporal uncertainty was introduced with a randomly variable interval (500 to 1000 ms) between the fixation point onset and the target onset, in order to discourage anticipatory responses.

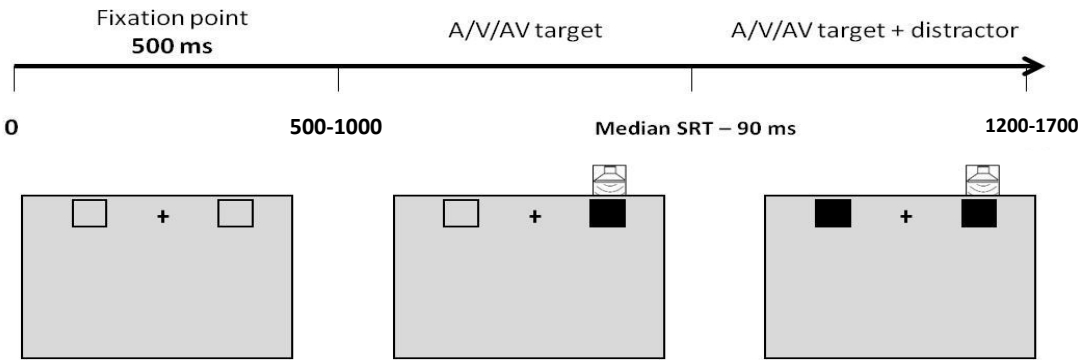


Figure 5.13 Experiment 9, stimuli and timetable of a typical trial. A schematic representation of a typical experimental distractor trial with the timeline (not in scale) in Experiment 9. Each trial started with a fixation cross together with two outline squares on the right and left upper corner of the screen, which lasted for 500 ms, followed by a random interval (500-1000). The target was a black square on a grey background, appearing always on the right side of the fixation cross and lasting 700 ms. The distractor was a black square presented always on the left side of the fixation point.

Participants performed a preliminary session with the three target conditions (A, V and AV) in order to compute the median SRT for each subject individually, which was used to determine the three different distractor onsets. This procedure was chosen in accordance with the Saccadic Inhibition hypothesis (Reingold & Stampe, 2002), which states that the maximum effect of the distractor is evident 90 ms after the saccade is initiated. Since we assume that different target types will result in different SRT distributions, three different distractor onsets were derived. In this way, we can be confident that any difference in the SI dip is due on factors other than the timing of the distribution, i.e. stronger resistance of certain distributions etc.

Participants were required to generate a single saccade toward the target (T) as soon as it appeared on the right side of the screen in the target only conditions, and to ignore the visual distractor in the distractor conditions.

To provide sufficient observations for a saccadic inhibition analysis, 150 trials per condition were presented (total number of trials = 900), divided into 4 experimental sessions, each composed of 9 blocks of 25 trials. The overall duration of the experiment was ~110 minutes.

Data screening

Saccades directed leftward (1.5%), saccades of less than 1° amplitude (3.2%), and saccades with a latency below 70 ms (5.1%) or above 500 ms (6.4%) were excluded. There were no difference in the amplitude of the three target types ($F_{5,45}$= 1.5, p= 0.05).

5.3.4.2 Results

SRT analysis

For each participant, the mean SRT was taken as the measure of central tendency for each condition. The distribution of mean values for each condition of the Distractor (2) x Target (3) is shown Figure 5.15.

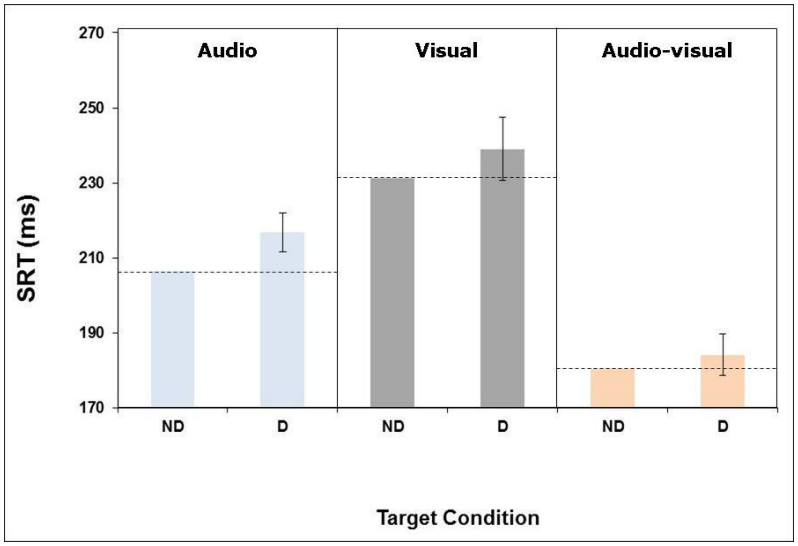


Figure 5.14. Mean SRT (ms) for Distractor (D) and Non Distractor (ND) condition, for the three types of targets (A, V and AV) in Experiment 9. In each target condition, the distractor cost is represented as the as the difference above the dashed line in comparison to each ND baseline (with the corresponding 95% confidence interval, calculated only for the difference between D and ND conditions, represented by the portion of the histogram above the dashed line).

In order to evaluate the effect of the target modality on the distractor cost, the analysis was performed on the difference values between corresponding D and ND conditions, for the three target modalities (Figure 5.14). Given the strong predictions I had about the results, a planned comparisons analysis was performed, which revealed a difference between A and AV ($F_{1,10}$= 6.4, p< 0.05), but no difference between V and AV ($F_{1,10}$= 1.1, p= 0.3). One-tailed t-tests against zero revealed a significant distractor cost for A and V conditions [A= 10 ms, t(10)= 4.3, p< 0.01; V= 8 ms, t(10)= 2.2, p< 0.05] but not for the AV condition [4 ms, t(10)= 1.7, p= 0.1].

The MRE in D and ND conditions was also investigated. The % MRE was calculated according to the previously-used formula: %MRE= 100*(min(A,V)– AV)/(min(A,V)) (Multisensory index, Stein et. al, 2009).

In this way, the percentage of enhancement in relation to the most effective unisensory stimulus (A or V) was calculated for every participant, in D and ND conditions. The prediction for this variable was that the %MRE would be weaker in the ND than in the D condition. Both %MREs were significantly different from zero [D= 14.0 % (SD= 7.6), $t(10)= 4.5$, $p< 0.001$; ND= 12.5% (SD= 7.5), $t(10)= 5.0$, $p< 0.001$]. Contrary to our predictions, the results showed that the %MRE in the D condition was stronger that the %MRE in the ND condition [$t(10)= 4.5$, $p< 0.001$]. This result is not surprising if we consider that there is no difference between AV and AVD conditions [181 vs. 184 ms respectively, $t(10)= 1.7$, $p= 0.1$]. Therefore, the difference in %MRE can be explained by a floor effect: AV target produced the same optimal speed of response in both the D and ND condition, and the difference in MRE was due to the difference in the two baselines (A/V and AD/VD conditions). This result tells us, at least on the basis of the SRT analysis, that an AV target produces the same response independently of the presence of a competing distractor. This conclusion has to be, however, further explored and confirmed by the Race model analysis.

Race model analysis

In order to investigate if the MI occurs in the AV conditions (distractor and non-distractor), two Race model predictions were derived for every single subject in the two different distractor conditions, and corresponding cumulative distributions for the A, V and AV conditions in the D and ND conditions separately (Figure 5.15 a). Unexpectedly, only 4 out of 11 participants showed a positive area of violation. For the remaining 7 participants, although the SRTs were faster in the AV conditions, the data did not violate the upper boundary predicted by the model.

In order to test the violation of the model in D and ND conditions, the Race model analysis was performed in the following way: SRT corresponding to every percentile separately was derived for the AV and AVD conditions and for the D and ND race model, for every subject separately. A series of t test was calculated on SRT of every percentile to

test the difference between the two AV conditions and the two models. As can be noted from Figure 5.15 b, there was no violation of the model in either AV conditions, and no differences between the conditions were observed (see the APPENDIX of the present thesis for a detailed explanation of the procedure).

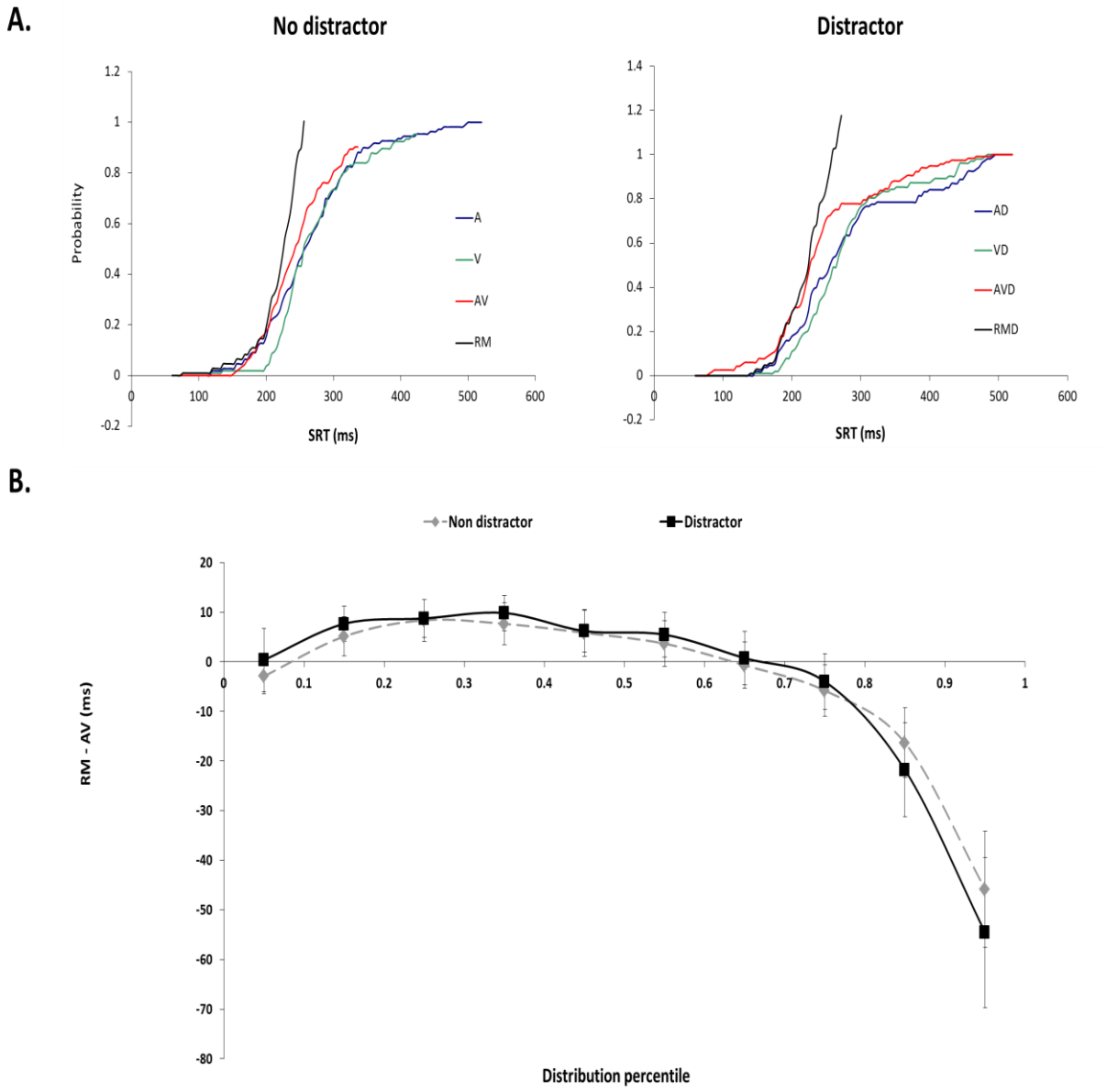


Figure 5.15 Experiment 9, Race model analysis. A) Cumulative distributions of a representative participant in D and ND conditions. Only 4 out of 10 participants presented a positive violation of the model. B) Group difference profile for the two AV conditions (D and ND) in relation to the corresponding Race models for every percentile. No significant violations of the model were observed in the two distributions and no differences between two profiles were observed.

SI analysis

Difference histograms were derived by bin-by-bin subtraction of the baseline histogram (A, V or AV target only condition) from the corresponding distractor condition histogram (AD, VD or AVD). For illustrative purposes, difference histograms were aligned to distractor onset and then averaged across participants for every distractor condition (Figure 5.16).

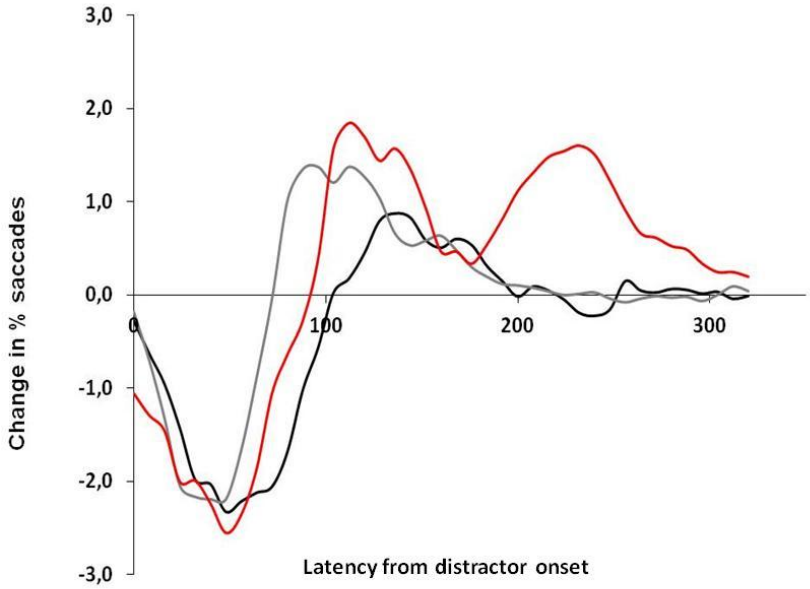


Figure 5.16 Experiment 9, Saccadic inhibition dip. Difference histogram (distribution dip) for each of the target conditions derived from the bin-per-bin subtraction of each Target only histogram (TA, TV or TAV) from the correspondent distractor histograms (A, V or AV).

Table.3. Measures of saccadic inhibition in Experiment 9 for the three target types

Target type	Magnitude (%)	Lmax (ms)	Duration (ms)	Recovery time (ms)
V	-3.9 (2.3)	70.5 (35.6)	90.0 (49.6)	106.9 (57.2)
A	-4.8 (3.8)	60.0 (21.2)	73.6 (55.4)	94.4 (50.9)
AV	-4.3 (2.7)	53.5 (13.9)	43.6 (29.9)	66.7 (24.5

Parameters from Table 3 were extracted for each difference histogram separately for each participant and entered into a within-subject ANOVA. Overall, the only parameter that differentiated the three dip profiles was the *Duration* ($F_{2,20}$= 8.48, p< 0.002). The post-hoc analysis revealed a difference between V and AV target condition (t(10)= 6.12, p< 0.001), with the AV dip having a shorter duration in comparison to the V dip. Overall, the dip profiles did not differ from one-another for dip magnitude ($F_{2,20}$= 0.62, p< 0.2) , Lmax ($F_{2,20}$= 2.79, p< 0.09), or recovery time ($F_{2,20}$= 1.71, p< 0.2).

5.3.4.3 Conclusions

The main aim of the Experiment 9 was to explore the relation between two different phenomena of the saccadic movements: the inhibition caused by a distractor (measured either as SI or RDE) and MRE. It has been demonstrated that both the inhibition and MRE can occur in the same experimental design. However, no clear interaction has been found between them. In other words, the MRE occurs with AV stimuli in relation to the unimodal A and V saccades regardless of the presence of a contralateral distractor. When we calculate the magnitude of the %MRE in relation to the unisensory baselines, we can see that it is stronger in the distractor vs. non distractor condition. This result is ambiguous and could be explained by the floor effect present in the AV condition: both the AV targets (D and ND) drove the response to an optimal level, preventing the inhibiting influence of the distractor to appear. This result could be interpreted in favour of a stronger resilience of the AV driven saccades toward distractors.

However, some cautions are needed when drawing this conclusion since we did not obtain a violation of the Race model, and the superiority of both AV conditions is in this case a statistical artefact rather that a genuine interaction, and therefore we cannot conclude that the multisensoriality conferred major resistance toward the distractor (Raab, 1982). This conclusion is further supported by the rest of the statistical analysis that was performed. For example, the general analysis on the Distractor cost did not reveal a significant difference between the A, V and AV target conditions (presumably due to a lack

of power in the main ANOVA), a trend in the data can be described on the basis of the evidence that only the visually and auditorily-driven saccades exhibited a DC different from zero, whereas no difference was evident in the AV target condition.

Overall, since no difference in the SI dip profiles has been noted we could conclude that the nature of the saccade did not affect susceptibility to distraction but the specific influence of MI could not be studied because not enough participants showed violation of the Race model.

As such, the lack of MI in AV conditions represents the main methodological imperfection of Experiment 9. A possible factor that explains this result could be the automaticity of the saccades to the target, which was always presented on the right. The high predictability of the target facilitated both responses to unisensory (A and V) and multisensory (AV) targets. In such a situation, the information from a unisensory target is already able to support an optimal level of response, so further benefit from an additional source of information (i.e. a second stimulus) is not significantly visible. This weakness of the experimental design should be remedied in the next experiments.

5.3.5 Experiment 10

In order to avoid the automaticity of responses present in Experiment 9, in Experiment 10 spatial uncertainty of the target was introduced. Presenting the target on the left or right of the fixation point, however, required a location for the distractor that would have to be different from the position of the target. For this reason, and in order to maximise the distraction effect, in Experiment 10 a generalized flash transient (similar to Experiment 8) was used instead of a lateralized square. The research questions addressed was similar to that in Experiment 8, but particular attention was dedicated to eliciting a strong MI in AV conditions.

5.3.5.1 Method

Participants

A group of 9 participants was tested (6 female and 3 male, age range 20-26). All had normal hearing and normal or corrected-to-normal visual acuity, gave their informed consent prior to the beginning of the experiment and received course credits or were paid for their participation.

Apparatus and stimuli

A fixation cross 0.5° x 0.5° was presented to the participant at the centre of the screen, with two empty squares (1.65° x 1.65°) placed on the right and left of the fixation cross. The empty squares were needed as placeholders in audio target conditions and in addition indicated to the participant the position of the appearance of the visual target. The distance between the target and the fixation cross was 7° of visual angle. The visual target was the same as in Experiment 2: a black square (0.3° x 0.3°of visual angle), presented at the centre of the empty square on the right side of the fixation cross (Figure 5.17). The auditory stimulus was a 100 ms white noise burst presented from a loudspeaker situated on the left or on the right side of the screen. The audio-visual target was determined by the simultaneous appearance of the visual and auditory target (both placed in the vertical centre of the two screen edges, with a horizontal displacement of 5° of visual angle between the centre of the visual target and the centre of the loudspeaker). The same distractor as in Experiment 8, a bright flash (25.6°x 4.2°) displayed for 30 ms at the bottom and the top of the screen, was presented across all the distractor conditions, regardless of the modality of the target. The onset of the visual distractor was different in the three target conditions, and was calculated by subtracting 90 ms from the average median SRT for each target condition, taken from the preliminary experiment. The duration of the target was 700 ms (Figure 5.17).

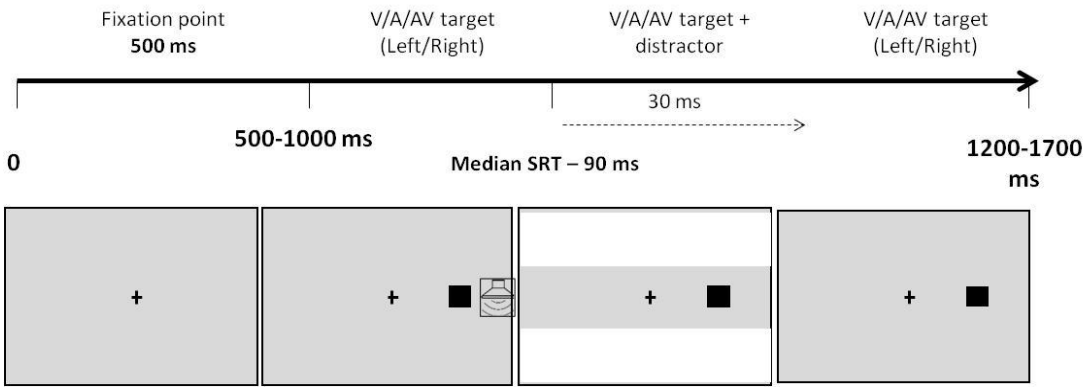


Figure 5.17 Experiment 10, stimuli and timetable of a typical trial. A schematic representation of a typical experimental distractor trial with the timeline (not in scale) in Experiment 10. Each trial started with a fixation cross, with a variable duration (500-1000 ms), followed by the target (a black square on a grey background), appearing randomly on the left or right side of the fixation cross and lasting for 700 ms. The auditory target was a white noise burst, originating randomly from two loudspeakers positionated on the right or left side of the screen. The AV target consisted of V and A targets presented simultaneously. The visual distractor consisted of a white flashing zone on the topo and bottom of the screen, lasting for 30 ms.

Procedure

The experimental design consisted of six experimental conditions, presented in a random order: visual target only condition (TV), in which the visual targets were presented randomly on the left or on the right of the fixation cross; auditory target only condition (TA), in which the white noise burst was emitted randomly from the left or right loudspeaker; the audio-visual target only condition (TAV), in which the visual and auditory target were presented simultaneously on the left or on the right from the fixation point, always spatially coincident; and the three distractor conditions (VD, AD, AVD) in which each target condition was presented with a visual flash transient appearing on the top and bottom of the screen.

At the beginning of each trial, participants were required to fixate the fixation cross in the centre of the screen. The duration on the fixation cross was variable and ranged from 500 to 1000 ms. Participants performed a preliminary session with the three target

conditions (A, V and AV) in order to compute the median SRT for each participant, which was used for the derivation of the three different distractor onsets (similarly to Experiment 9).

Participants were required to generate a single saccade toward the target (T) as soon as it appeared on the left or right side of the screen in target only conditions for three type of targets: auditory (A), visual (V) and audio-visual (AV).

To provide sufficient trials for a saccadic inhibition analysis, 180 trials per condition were presented (total number of trials= 1080), divided into 3 experimental sessions, each composed of 10 blocks of 36 trials (6 trials per condition). The overall duration of the experiment was 130 minutes.

Data screening

Saccades directed in the direction opposite to the target (3.5%), saccades of less than 1° amplitude (2.3%), and saccades with latency below 70 ms (3.1%) or above 500 ms (2.1%) were excluded.

The amplitude of the three types of saccades (A, V and AV) were also taken into account and entered into a two way ANOVA with Target (A, V and AV) and distractor (D, ND) as main factors. A main effect of the factor target was present ($F_{2,16}$= 3.75, p< 0.05), while no effects of the factor distractor ($F_{1,8}$< 1) or of the Target x Distractor interaction were found ($F_{2,16}$< 1). The significant effect of the target was due to the larger amplitude of the auditory guided saccades (4.84°) in comparison to the visual (4.61°) or AV-guided saccades (4.61°). Because of the vertical distance between the visual target and the position of the loudspeakers (5°), the auditorily-driven saccade may have been attracted by the loudspeaker (although hiden) rather than by the placeholder, influencing the amplitude of the auditorily - driven saccade.

5.3.5.2 Results

RDE analysis

The median SRTs were entered into a Position (2) x Distractor (2) x Target (3) within-subject ANOVA. Figure 5.18 shows the mean of medians in each condition.

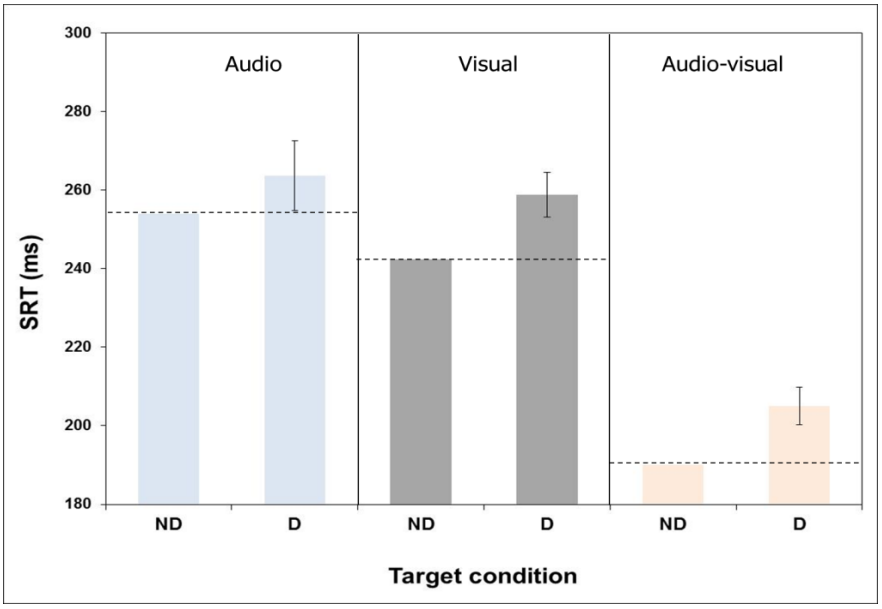


Figure 5.18 SRT (ms) for distractor (D) and Non distractor (ND) condition, for the three types of targets (A,V and AV) in Experiment 10. Similarly to Experiment 9, the Distractor cost is represented by the area above the dashed line in comparison to each ND baseline (with the corresponding 95% confidence interval, calculated only for the difference between the ND and D condition, represented above the dashed line).

Similarly to Experiment 9, the analysis was performed on the difference values between each D and ND condition, for the three target modalities (Figure 5.18). Contrary to Experiment 9, all the differences in Experiment 10 were significant against zero: A= 10 ms, $t(10)= 2.5$, $p< 0,03$; V= 17 ms, $t(10)= 5.7$, $p< 0.001$; and AV= 15 ms, $t(10)= 6.3$, $p< 0.001$. No significant differences were found between the distractor costs in the different conditions ($F(2,16)= 1.5$, $p< 0.2$).

Contrary to Experiment 9, no difference in the % MRE were found in D and ND condition (15.7% vs. 16.7% respectively, $t(8)= 1.5$, $p< 0.2$).

Race model analysis

The Race model analysis was performed on the data similarly to Experiment 9. In Experiment 10, the majority of the participants (7 out of 9) revealed a violation of the model, in both distractor and Non distractor conditions (Figure 5.19 a). The areas of violation were calculated (see APPENDIX) for the D (5.2) and ND (7.6) conditions, and the differences were statistically significant ($t(6)= 3.4$, $p< 0.02$), suggesting that the strength of inhibition was weaker when a distractor was present. In particular, the occurrence of the distractor affected the time at which the cumulative distributions first violated the Race model prediction: distractor (224 ms) vs. no-distractor (237 ms), $t(6)= 3.8$, $p< 0.01$. Figure 5.19 b represents the difference between the AV and AVD conditions in relation to the corresponding ND and D Race model (see APPENDIX for the explanation of the procedure). As can be seen from the graphic, a strong violation of the model, present in a large number of percentiles, is evident in the ND condition, whereas the D condition violated the model only in 1 percentile of the cumulative distribution.

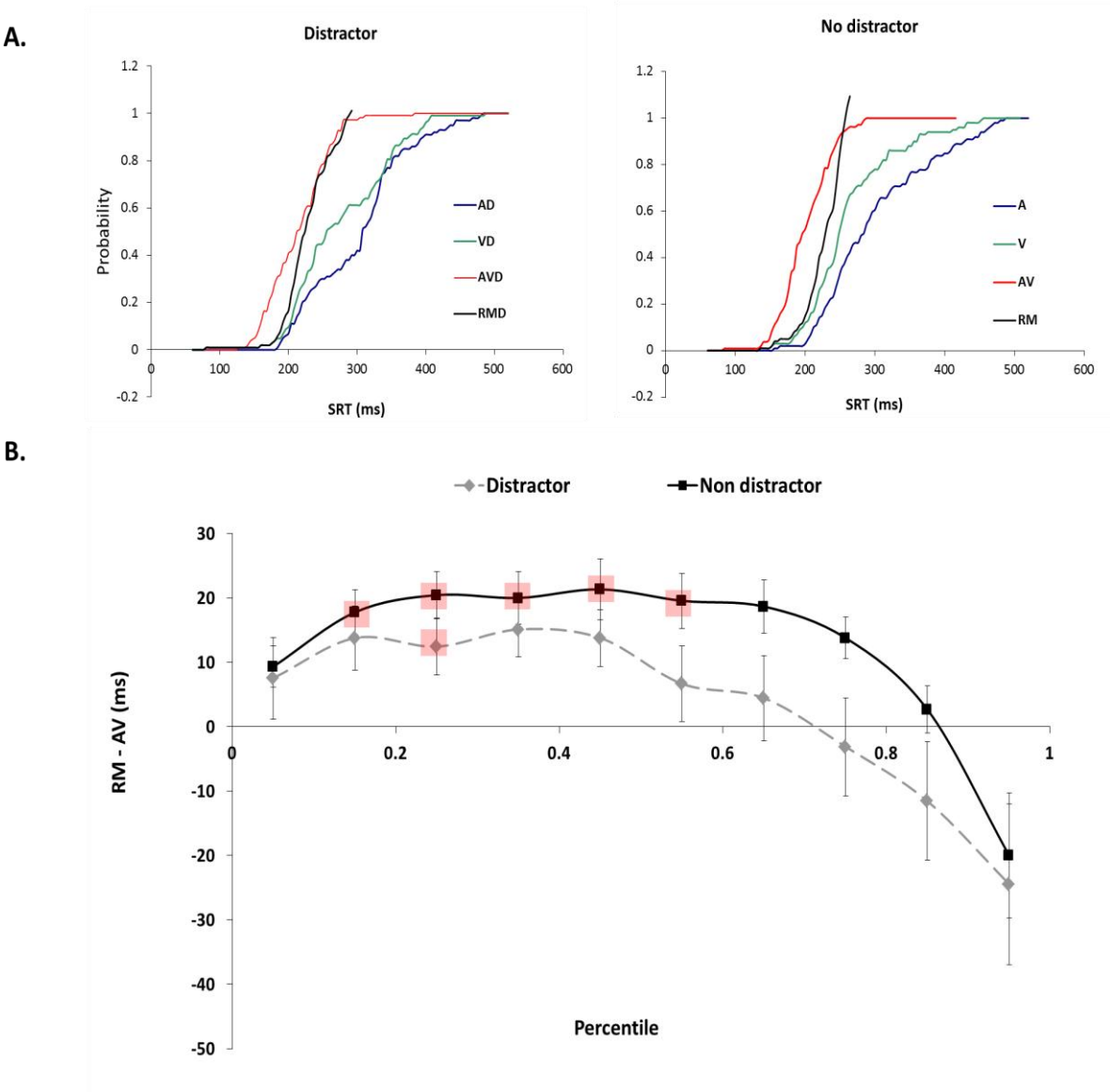


Figure 5.19 Experiment 10, Race model analysis. A) Cumulative distributions of a representative subject in the D and ND condition. B) Group difference profile for the two AV condition (D and ND) in relation to the corresponding Race models for every percentile. Red squares indicate the percentiles that violated the Race model. The ND condition exhibited a strong violation of the model in five percentiles, while the distractor condition exhibited only a small violation in one percentile.

Saccadic inhibition analysis

The analysis of the dip parameters (Table 4) revealed a similar pattern to the one observed in Experiment 10. The only significant differences were observed for the parameter *Duration* ($F_{2,16}$= 9.43, p< 0.01), with the auditory target dip exhibiting the shortest duration, both in comparison to the V (t(8)= 3.49, p< 0.01) and to the AV (t(8)= 3.54, p< 0.01) target condition, as revealed by the post-hoc analysis.

Table.4. Measures of saccadic inhibition in Experiment 9 for the three target types.

Target type	Magnitude (%)	Lmax (ms)	Duration (ms)	Recovery time (ms)
V	-4.5 (1.7)	62.2 (13.1)	120.0 (22.1)	72.0 (19.3)
A	-4.9 (1.8)	76.4 (19.3)	68.0 (28.3)	76.4 (23.9)
AV	-6.1 (1.1)	57.8 (14.5)	124.4 (24.2)	58.7 (9.5)

Apart from the duration, the three dip did not differ in the *Magnitude* (A $F_{2,16}$= 1.58, p< 0.2), in the *Lmax* (A $F_{2,16}$= 2.28, p< 0.1) or in the *Recovery time* (A $F_{2,16}$= 1.48, p< 0.2).

5.3.5.3 Conclusions

In Experiment 10, the introduction of spatial uncertainty over target presentation (A,V and AV) seemed to facilitate the occurrence of MI. SRTs in both AV and AVD conditions were faster in comparison to the faster US target, as revealed by the %MRE. Moreover, the %MRE did not differ in the D and ND condition. The analysis of the RDE for the three target condition revealed similar amounts of RDE in the three conditions, contrary to our expectations.

The analysis of SRT did not reveal any difference between the distractor and non-distractor condition, indicating that the multisensoriality of the target did not confer any

additional resistance toward the competing distractor. On the contrary, the presence of a distractor resulted in the same “distracting” cost across all target conditions.

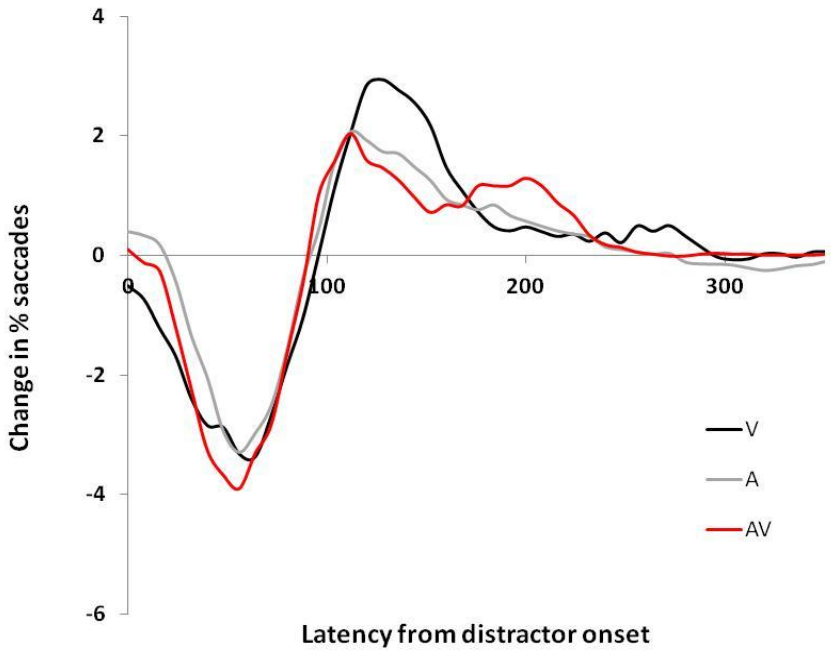


Figure 5.20, **Experiment 10, Saccadic inhibition dip**. Difference histogram (distribution dip) for each of the target conditions derived from the bin-per-bin subtraction of each Target only histogram (TA, TV or TAV) from the correspondent distractor histograms (A, V or AV).

The hypothesis of the major influence of the distractor has been confirmed also by the Race model analysis, were the ND condition showed a strong violation of the model, whereas almost no violation was present in the D multisensory condition.

For the SI analysis, we can assume that, regardless of the target modality, it is the timing of the distractor that determines the strength of the inhibition. This assumption is in line with the Saccadic Inhibition hypothesis, which states that the key factor that determines the strength of the inhibition is the time delay between the distractor onset and the planned saccade. Following this hypothesis, three different distractor onsets were linked to three different SRT distributions for the three target condition. This means that,

regardless of the target modality, the visual distractor that appeared always 90 ms after the target onsets and produced the same dip in the three conditions.

In conclusion, even if the main prediction of the major resistance of AV guided saccades toward a distractor were not proved, the data highlighted other aspects of how MI (of bimodal targets) and concurring distractors interact, such as the fact that a simultaneous visual distractor can interfere with the process of multisensory integration preventing the integration to occur.

5.4 General conclusion

The SC is the main structure in the process of guiding orientation behaviour by translating sensory signals into appropriate motor commands. Multiple competing stimuli are constantly present in our environment, and the SC has the role to select one target among them. Two main effects have been studied in the literature regarding target selection in the SC. The first is MI, showing how multimodal spatio-temporally coincident stimuli lead to an enhanced activity at both neuronal (Stein & Meredith, 1993) and behavioural (Todd, 1912) levels. On the contrary, other studies have focused their attention on the competing and inhibiting effects caused by multiple targets, describing the circumstances that cause the RDE and SI. Contrary to MI, these phenomena have been studied mainly in the visual domain, related to the fact that an unexpected onset of the distracting stimulus is thought to elicit a new motor program or a shift of attention interfering with the on-going program elicited by the target of the saccade.

The two main goals of this chapter were to explore the multisensoriality of the RDE and SI, and to investigate the behavioural consequences of a situation involving both multisensory target and multisensory distractors, and the influence that they have on the saccade generation. In particular, it was predicted that the multisensoriality would confer more strength both to the distractor (potentiating the distracting effect) (Experiments 7 and 8), and to the target (giving more strength to resist a competing distractor; Experiments 9 and 10).

Experiments 7 and 8 aimed to highlight the influence of auditory and audio-visual distractors on a visually guided saccade, compared to a visual distractor. The data showed that generally an AV distractor produced a stronger inhibition (on the basis of SRT analysis), whereas no RDE was obtained with A and V distractors (due to methodological imperfections). However, the A, V and AV distractors produced a SI in the distribution, proving the multisensoriality of the phenomenon. A similar result was obtained in Experiment 8, in which generalized visual and auditory distractors were used instead of lateralized distractors in Experiment 7. Given the stronger saliency of the distractors in Experiment 8, a stronger RDE was obtained overall with all the distractors. The superiority

of the AV distracting effect over A and V distractors was demonstrated both in the SRT analysis and in the SI analysis. However, the AV distractor superiority did not violate the Race model of integration. These results are, however, in line with the results obtained by Diederich and Colonius (2007). The authors presented an auditory or audio-somatosensory contralateral distractor at 50 ms after the target onset. In their study, a significant inhibiting effect was observed with the bimodal audio-somatosensory distractor, but no effect with the auditory distractor alone (similarly to Experiment 7 of the present chapter). A possible lack of the distracting effect in their experiment may be due to the early distractor onset (50 ms) presented after the target onset, whereas in Experiment 8 of the present chapter the auditory distractor appeared on average 85 ms after the target onset, producing a stronger inhibiting effect. The data are in line with the TWIN model (Diederich & Colonius, 2007), which states that the enhancing effect of an accessory stimulus is achieved only if it anticipates the target. Since in our studies the auditory stimulus always appeared after the target onset, no warning effect was present in the results.

Experiments 9 and 10 addressed the question of whether a multisensory target confers stronger resistance toward a visual distractor. In Experiment 9 lateralized targets (A,V and AV) and distractors were presented. The overall inability to elicit the MI (i.e. violation of the race model) in both AV conditions precluded a clear answer to the research questions. In Experiment 10, the introduction of spatial uncertainty set a better basis for the manifestation of the MI (indexed by violations of the Race model) in AV conditions. On the basis of the SRT analysis, both the MRE in D and ND conditions were comparable. Contrary to Experiment 9, in this case the AV and AVD condition were significantly different, with the AV responses being faster that AVD responses. This means that, according to the previously adopted logic, although the magnitude of the enhancement in relation to the baseline is comparable, the resulting AV response were better in the ND than in the D condition, suggesting that in this case the multisensoriality did not offer a stronger resistance to the competing distractor. This result is confirmed also by the Race model analysis, in which the model is highly violated in the ND condition, whereas only a weak violation (in one percentile over 10) was present in the D condition. We can speculate about this difference in the result in Experiments 9 and 10: the spatial

uncertainty of the target facilitated the MI effect, whereas a generalized flashing distractor (and therefore more salient than the target) had a stronger effect in comparison to the lateralized distractor in Experiment 9.

To sum up, the answer to my initial question would be the following: an AV distractor does produce a stronger inhibition (Experiment 8), but an AV target does not confer more resistance toward a competing distractor (Experiment 10).

The results, although confirming only partially my initial hypothesis, are in line with the *temporal account* of the saccadic inhibition proposed by Born and Kerzel (2008, 2011) and Bompas and Sumner (2009b), and with the *Saccadic inhibition hypothesis* proposed by Reingold and Stampe (2002). The study by Born and Kerzel (2008) described in previous paragraph already pointed out that the temporal overlapping of the target and distractor represents a crucial factor in determining the strength of the RDE. In their *temporal account* of the saccadic inhibition they suggest that the contrast of the target or distractor modulates the arrival time of the corresponding signal in the oculomotor system: high-contrast stimuli provoke an earlier response in the oculomotor structures than low-contrast stimuli. If we assume the existence of competitive processes between target and distractor elicited activities, then we can conclude that a distractor inhibition will be the strongest when distractor- and target-related activity will maximally overlap in time. In their later study, Born and Kerzel (2001) introduced various SAO between the target and distractor, in order to further investigate this hypothesis. Their result confirmed their prediction: the lower is the distractor contrast with respect to the target contrast, the earlier the distractor has to be presented to produce a strong RDE, proving that varying target and distractor characteristics modulates at which point in time the respective stimulus signal reaches the oculomotor structures. A strong RDE can only be observed when target and distractor signals show large temporal overlap in the oculomotor structures. This conclusion confirmed previous findings from Bompas and Sumner (2009) that used peripheral distractors and showed that for a target of a given contrast, the RDE is strongest when the SOA is shifted according to distractor contrast.

My data are in line with the reported studies and complements their results in the following way: in order to ensure that the activity of the distractor and the one of the

target will overlap in time (and to compensate for the different processing time needed in the case of different saliencies of distractor and target), the distractor was time-locked to the expected time of saccade initiation at a per-participant level. In this way, independently of the saliency (or identity) of the target –A,V or AV- the inhibition will always occur at the same time. Moreover, according to the SIH (Reingold & Stampe, 2002), the onset of the distractor occurred 90 ms after the median SRT, when its inhibiting influence is expected to be maximal. The results confirmed that, if we compensate for the temporal delay due to the different processing time, the effect of a A and V distractors on the saccade is always the same (the dip of the inhibition does not change with A, V or AV targets). However, this was not the case with AV distractors, who did produce a stronger dip in the distribution, even when time-locked to the distribution. This is also in line with the result of Born and Kerzel (2011), who also reported that, at optimal SOA, the magnitude of inhibition was still influenced by the specific characteristics of the distractor and of the target. Particularly, the inhibition increased with increasing distractor contrast, suggesting that a stronger signal perturbed saccade programming more than a weak signal. Interestingly, the inhibition was less influenced by target contrast, similarly to my results. The authors suggested that the modulation of the characteristics of the distractor influences a mechanism that is different from the one involved in the processing of the target, and therefore it may be not legitimated to expect the same result in the two cases. Moreover, as the authors further suggested, target characteristics may be modulated through top-down influences on the target signal.

Overall, my data are in line with the studies here described, and seem to suggest that the characteristics of the stimuli, especially when they act as distractors, do not influence only the time needed for the signal to arrive to the saccade map, but also the subsequent motor competition processes.

In conclusion, the experiments presented in this chapter can help shed some light on the processes of target selection in saccade generation in a multisensory environment. Further research is, however, necessary to complete the range of possible target-distractor combinations. Some further speculations about the data will be given in the General conclusion chapter.

CHAPTER 6:

GENERAL DISCUSSION

Chapter 6:
General discussion

The general topic of the present thesis regards the multisensory integration (MI) of auditory and visual stimuli. MI is a growing research field that goes beyond the traditional "sense-by-sense" approach, and emphasizes the synergy among the senses and the combination of their information. Two effects of MI have been described in the literature: response improvement in classification and detection tasks (revealed by an increase of response speed and/or accuracy) and multisensory illusions (see Chapter 1).

In the present thesis MI has been assessed by evaluating the effectiveness of multisensory stimuli, relative to the composing unisensory stimuli, in evoking some type of response in the organism. Such a gain in effectiveness is generally named *multisensory response enhancement* (MRE), although other terms are also available in the literature (e.g., redundancy gain, redundant target effect, etc.). As such, the MRE has been the principal dependent variable across the chapters of the present thesis, although other variables have been taken into consideration such as the *distractor cost* in Chapter 5 (defined as the increase of the SRT when a distractor is present).

The research undertaken in this thesis primarily aims at extending the current understanding of the influence of multisensory audio-visual integration on human behaviour. In particular, I used a simple detection task to study the influence of various factors on manual responses (spatial coincidence in Experiments 1-3, shape congruency in Experiments 4-6), as well as the influence of distractors on the generation of saccades in Experiments 7-10.

The role of attention represents a secondary *motif* of my thesis. In my experiments I used either the focused attention paradigm (with attention focused on vision) or the redundant target paradigm (with attention distributed over auditory and visual modalities). Although I did not run a specific experiment involving the direct manipulation of attention, I will draw some conclusions about the role of attention in MI based on the

comparison of different experiments presented in my thesis, as well as on the comparison of my results and those obtained in other studies

6.1 Summary of experiments and conclusions

In general, my thesis concerns audio-visual integration. Specific experiments cover three sub-topics: the spatial rule mediated by the Superior Colliculus (SC; Chapter 2); the influence of audio-visual shape congruency in a focused attention detection task (Chapter 3) and in a redundant target detection task (Chapter 4); and the influence of audio-visual interaction on the *distractor effect* studied in the context of the generation of saccades (Chapter 5). This chapter includes: a summary of the main findings from each experiment presented in Chapters 3-5; general conclusions in light of the evidence reviewed in Chapter 1, so as to offer a global view of the role of audio-visual interactions in manual and saccadic responses; additional comments on my results in light of the data present in literature.

6.1.1 The spatial rule mediated by the Superior Colliculus

To provide converging evidence on the role of the SC in MI I ran three experiments in which two different methods were adopted for isolating the chromatic component of SC-ineffective (blue) targets from their luminance: the random modulation technique in Experiments 1 and 2, and the tritanopic technique in Experiment 3.

In Experiments 2 and 3 I replicated results consistent with the spatial rule of MI, finding a superiority of MRE when visual and auditory stimuli were spatially congruent. The superiority, however, was evident only with stimuli that supposedly activated the SC, or at least the magnocellular pathway which provides an input to the SC. In the condition in which the SC was not activated (with S-cone stimuli and random luminance technique in Experiment 2 and with the tritanopic stimuli in Experiment 3), no superiority of spatially coincident vs. spatially non coincident stimuli was obtained.

In Experiments 1-3 I used a focused attention paradigm (attention focused only on the visual modality) to replicate results from studies that adopted a redundant target paradigm (attention distributed over auditory and visual modalities), like the one by Leo et al. (2008).

To evaluate and compare the MRE amounts in different conditions I always considered a relative measure (the %MRE, Stein et al., 2009) in which the speed of response to the homologous target in unisensory trials was taken as a baseline, for each individual observer.

To evaluate the possible differences between the redundant target paradigm utilized by Leo et al. (2008) and the focused attention paradigm utilized in my experiments, let me consider the pattern of response times and % MRE amounts in each study:

- *US conditions*: average RTs in the unisensory condition were similar for the redundant target paradigm and the focused attention paradigm. Admitedly, this is just a speculative assumption, since I am not able to verify the statistical valence of such a conclusion.

- *MS conditions*: The RTs in the multisensory conditions were overall longer in my Experiment 2, compared to Leo et al. (2008). The reason for such a discrepancy is unknown and further experiments would be needed to clarify this point. In the absence of any additional result, we can speculate about a potential role of the paradigm adopted (which constitutes the major difference between the two studies) in determining this difference.

- *Pattern of data*: except for the overall weaker effect, data from my Experiment 2 follow the same pattern as those by Leo et al. (2008), who found a congruency advantage for SC-effective but not SC-ineffective targets.

As a general conclusion, data from focused attention and redundant target paradigms display a similar pattern, though they differ regarding the overall amount of facilitation (which can be due to the different distribution of attention or other methodological differences).

Such a pattern is consistent with the TWIN model by Colonius and Diederich (2004), already described in Chapter 5 in relation to the saccadic generation. Colonius and Diederich (2004) stated about the focused attention task:

> *„[...] Focused Attention Task: When the task is to orient toward the target stimulus ignoring stimuli from other modalities, the first stage terminates when the target peripheral process ends. Multisensory integration occurs only if the nontarget stimulus wins the race in the first stage opening a time window such that the termination of the target peripheral process is enclosed in the window..."*

and about the redundant target task:

> *„[...]When the task is to orient toward the first stimulus detected no matter of which modality, the first stage duration is defined by the winner's peripheral processing time, and the window of integration is opened by whichever stimulus wins the race."*

If we assume that also in my experiments, as well as in those by Leo et al. (2008), the time needed to process the auditory stimulus was shorter than the time needed to process the visual stimulus, we can conclude that, regardless of the paradigm adopted, it was always the auditory stimulus to open the 'window of integration'. Moreover, we can speculate that the processing time for the auditory nontarget in our experiment was longer (since the participants had to inhibit the auditory response), therefore resulting in a weaker effect (in comparison to Leo et al., 2008).

Regarding the SC-ineffective stimuli, no superiority of spatial coincidence was found in my experiment. Assuming that SC-ineffective stimuli have been processed in the cortex, I suggest that the rules of MI do not apply to MI processes that might occur in the cortex. In fact, physiological evidence in the cat indicates that MI in cortical association areas (i.e., the anterior ectosylvian fissure and the lateral sulcus) is less influenced by the temporal and spatial congruency of multisensory stimuli (Wallace et al., 1992) and therefore may not directly mediate the behavioural consequences of MI.

In conclusion, despite the s-cone methodology has been widely adopted in the studies on audio-visual integration (Sumner, Adamjee, & Mollon, 2002), and the assumption of the distinction between SC-effective and SC-ineffective stimuli is taken as clear-cut, there are some aspects that should be further considered. For example, it is often neglected that even the SC-effective stimuli are lickely to be elaborated by the cortex, eventually in a later stage of the processing. Therefore, the effects that are observed with „SC-effective“ stimuli cannot be taken as a pure result of the influence of the SC.

6.1.2 Cross-modal shape congruency

The experiments in Chapters 3 and 4 were concentrated around one main topic: the congruency of an intrinsic property (i.e., shape) of auditory and visual stimuli. The experiments in Chapter 3 utilised a focused attention paradigm, while the experiment in Chapter 4 utilised a redundant target paradigm; this manipulation allows us to speculate about the role of attention, although additional experiments are needed.

An important aspect of experiments in Chapters 3 and 4 regards the use of *shape* of visual and auditory stimuli as the crucial variable. Parise and Spence (2009, Exp. 2) found an effect of cross-modal shape congruency in temporal order judgement task. My experiments represent a novel contribution since, taken together, they constitute the first study showing that detection of visual and auditory stimuli can benefit from the cross-modal correspondence of a complex attribute. So far, most studies utilised simple dimensional features that can be easily mapped on the quantitative continuum of subjective intensity, on which stimulus changes can be expressed as an increase (e.g., brightening) or decrease (e.g., darkening). A higher level of physical energy, presumably coded at the neuronal level by an increase in the overall discharge rate (Recce, 1999), corresponds to an increase of sensory magnitude (brighter light, louder sound). In such cases, a correspondence between brightness and loudness is intuitive, since both are intensive dimensions.

The correspondence between visual and auditory *shape* is less intuitive. However, its effect on MI is in line with a vast research field that has been concentrated for years on the perception of shape or *contours* (as commonly referred in the literature), and aims to highlight the importance of contours (in visual, auditory or somatosensory modality) in the perception of objects and events.

As regards the auditory modality, it is of some interest that the auditory contour (defined as the acoustic sequence of different frequencies and intensities in contrast to pure tones) is among the first aspects of the acoustic world perceived by human neonates (Dowling, 1982). Patterson, Uppenkamp, Johnsrude and Griffiths (2002) have demonstrated that pitch patterns (defined as the combination of different pitches in a sequence) are represented in the primary auditory cortex including the posterior superior temporal gyrus and planum polare. McKenna, Diamond, & Weinberger (1984) reported evidence of a selective activation of single neurons in the auditory cortex of animals to tonal contours (obtained combining different frequencies in a sequence).

There has been even more studies reporting the importance of the shape (or contours or curvatures) in visual perception. For example, interestingly for the purpose of my experiments, Gallant, Braun, & Essen (1993) studied the selectivity for polar (circle, spiral- and star-patterns), hyperbolic (curvature patterns), and Cartesian (lines, edges, etc.) image patterns in some cells in area V4 in the macaque visual cortex, reporting that many cells are more sensitive to polar and hyperbolic patterns than to Cartesian patterns. Moreover, some cells were selectively sensitive to star-patterns but not to circle patterns, while others were sensitive to curvature around one direction but not to the other directions. These results are particularly interesting since the stimuli used by Gallant, Braun, & Essen (1993) are similar to the stimuli used in my experiment. Even more interestingly, Horridge (2000) has shown that honeybees can be taught to discriminate between circle and star patterns. His experiments also suggest that the bees have ’tangential’ and ’radial’ filters, i.e. cells that are sensitive to edges directed out from the center of the eye’s fixation point (e.g. edges in a star pattern) and edges directed orthogonally to those (e.g. edges in a circular pattern).

Another interesting study is by Humphreys et al. (1994), who studied the effects of lesions in the parietal lobe (Balint’s syndrom) and reported that these patients had

difficulties in recognizing squares represented by lines when another object represented by corners was present. On the other hand, when the square was represented by corners and the other object by lines their performance was normal. The authors suggested that this may indicates that corners outvote lines, and that they therefore are more important. This may also be because corners have a stronger influence than lines upon the attention mechanism.

Given this overview on the state of the research about shape perception, it seems that the contours, shape or curvatures have an important role in the perception of the environment. The study of how this information in combined across different modalities seems to be a logical step forward, and my data represent a first attempt.

I would summarize the results obtained in Experiments 4 and 5 (Chapter 3) and Experiment 6 (Chapter 4) around five main points:

I. *Shape congruency effect*

In agreement with Parise and Spence (2009) I have confirmed that a complex feature such as the shape of visual and auditory stimuli can lead to a congruency effect. This point represents a fundamental difference in relation to the past studies on CMC, which often studied *dimensional* features such as visual brightness, tactile roughness, or auditory loudness, etc.

II. *Early processing of cross-modal interactions*

The congruency effect measured in my experiments was obtained in an implicit task (speeded detection). The automaticity of such interactions is not so surprising if one assumes that they may be present from our early ages. For example, in the study by Meltzoff and Borton (1979) 1-month old infants were presented orally with a smooth or nubby pacifier. The authors reported that the infants looked longer at a matching shape than at a novel shape with which it was paired, proving that they can transfer information about the shape of objects from touch to vision. As already commented, these results suggest about an *early holistic processing* of the stimuli (Garner, 1974).

III. *Cross-modal congruency (CMC) and Inverse effectiveness rule (IER)*

Although my first hypothesis was that target effectiveness and cross-modal congruency (CMC) will operate in the same direction, with weaker targets resulting in a stronger MRE and a stronger congruency effect, this was not the case. Instead, the congruency effect seems to be stronger at higher contrast (supposedly because the information about visual shape is more efficient), while the total amount of MRE is stronger when contrast is lower (according to the IER). A further analysis of the data is given in paragraph 6.1.3.

IV. *Focused attention paradigm and redundant target paradigm*

The congruency effect can arise in both focused attention and redundant target paradigms. However, it is facilitated in the former one. When attention was distributed over visual and auditory modalities, the congruency effect occurred in both (tzk + spiky) and (wow + curvy) combinations, while it occurred only in the (tzk + spiky) combination in the focused attention paradigm. I suggest that the direction of attention facilitates the interaction between specific features of the visual and auditory stimuli, and prompt the congruency effect to arise.

V. *Multisensory response depression*

In Experiment 6, the incongruent combination between a curvy visual shape and a tzk (spiky) sound did not result in a super-additive MRE. In fact, the MRE did not differ from the best US response. Moreover, if we analyse data according to the race model, we can see that in the case of the tzk + curvy combination the race model boundary was not violated. This means that, when two stimuli are particularly dissimilar, they are treated as two different events. In fact, their interaction is better explained by the race model: the visual and auditory stimuli are processed separately, and the faster of the two trigger the response. This point is particularly interesting, because it suggests that the temporal and spatial coincidences alone are not sufficient for the integration to happen. We can

therefore assume the existence of a fourth rule of MI: beside the temporal, spatial and inverse effectiveness also the *congruency rule*.

What can these results tell us about audio-visual processing? In Chapter 1 I already described the two main models that tried to explain the interactions among multisensory information: The Race model and the co-activation model. In the light of my results in Experiment 6, I can exclude the Race model explanation, assuming that information from auditory and visual stimuli was combined in the activation of the response.

However, which kind of combination of auditory and visual information takes place is not so obvious. Miller (1991) proposed that the combination could be either a summation of independently produced activations or an interaction of the processes involved in detecting the redundant targets (independent vs. interactive co-activation). In the case of *independent co-activation*, the information about stimuli from both modalities is processed separately from each other. It is only in the final stage that the output of these processes is combined in the triggering of the response (Morton, 1969). Under this logic, we could expect that the characteristics of the stimuli are irrelevant since they do not interact before the final stage.

With *interactive co-activation*, the processing of information in one modality is affected by the presence and the identity of target information in the other modality. In particular, the final output of such interactions has some properties that will affect the triggering of the response. Under these assumptions, we can predict that pairs of stimuli that are more or less congruent (in relation to some intrinsic characteristics) will be more or less efficient in triggering the response (Long, 1976).

Together with the previous evidence from Miller (1991), my findings in Chapters 3 and 4 therefore lend support to the idea of an interactive co-activation model, which postulates that cross-modal interactions take place at a low level of sensory processing (Miller, 1991). According to this view, I suggest that targets contribute differentially to the total activation, depending on their identities (and on their reciprocal congruency). Miller (1991) has proposed that the congruency gain might depend on connections between the internal representations of different targets: i.e. the representation of one target facilitates

or triggers the representation of the other, by mean of facilitatory connections among them.

To sum up, a few considerations about MI can be raised from these conclusions. The interaction proved by the violation of the Race model prediction is based only on the energy detection level. However, the interaction demonstrated in my experiments and in those by Miller (1991) postulates an analysis of stimulus identities that cannot be explained by the simple energy detection. An extended model, which considers interactions at a more central level, would be therefore needed.

6.1.3 On the Inverse effectiveness rule

A special note will be dedicated to the Inverse effectiveness rule (IER), supported by my data. The IER states that the weaker is the response to US stimuli (expressed both at neuronal and/or behavioural level) the stronger is the integration between the US components in the MS condition. The rule has been widely investigated at a neuronal and behavioural level in cats (Alvarado, Vaughan, Stanford, & Stein, 2007; Alvarado, Stanford, Vaughan, & Stein, 2007; Stein, Huneycutt, & Meredith, 1988), in macaques (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Kayser, Petkov, & Logothetis, 2008), and in humans with neuroimaging (Stevenson & James, 2009) and behavioural techniques (see Homles, 2009 for a review).

As regards the study at the neuronal level, Perrault, Vaughan, Stein, & Wallace (2003) plotted the magnitudes of all multisensory interactions for a population of SC neurons as a function of the size of the best modality response, and reported that the response enhancement declined with the increasing in the magnitude of the dominant modality response. Similarly, in an fMRI study, Stevenson & James (2009) have explored whether the audio-visual speech perception and object recognition, usually mediated by the superior temporal sulcus (STS), also follows the IER. In the object recognition task, participants performed a two-alternative forced-choice decision of whether the object was a hammer or a paper cutter. In order to vary the level of difficulty of the task, the authors modulated the signal-to-noise ratio (SNR), corresponding to 55, 65, 75, 85 and 95% of accuracy. The same procedure was adopted with speech stimuli, which represented a

female actor pronouncing ten nouns. In both tasks, BOLD responses in STS were analysed according to the criteria of superadditivity (A+V< AV); and the authors reported that the AV responses exceeded the superadditive criterion at 55 and 65% of accuracy, with a less strong significance at 75% and no significant exceeding at higher SNR levels, claiming to have found evidence of the IER.

Beside the fact that single neuron and neuroimaging studies report clear and coherent evidence in favour of the IER, the results from behavioural studies in humans seem less clear, since some studies claimed to have found evidence supporting the rule (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003; Serino, Farnè, Rinaldesi, Haggard, & Làdavas, 2007), while others have reported contradictory results (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Behavioural experiments aim to replicate the results at neuronal level: the MI should be strongest in participants or conditions with the weakest responses or worst performance in unisensory conditions. I will here report a brief summary of the most representative ones. In a study by Hairston et al. (2003), the visual acuity of the participants was degraded using myopia-inducing lenses, which resulted in a 25% decline in visual performance. The performance of myopic participants was compared with the performance in condition of normal vision. With this methodology, the authors observed that in a visual localization task, participants with normal vision did not benefit from the addition of a spatially-coincident auditory cue. Conversely, the improvement under visual-auditory conditions during induced myopia resulted in performance that was equivalent to that seen for the same multisensory targets under conditions of normal viewing. In a later study, Bolognini et al. (2005) showed that combining two US stimuli enhances visual responses when a visual stimulus alone is not capable of evoking a response and thus below threshold.

However, beside the proofs in favour of the IER in behavioural tasks, some authors have suggested that the inverse effectiveness effect might be a statistical artefact rather than a genuine phenomenon. For example, Holmes (2007) argued that the choice of performance measure and statistical analysis may profoundly influence the likelihood of a given dataset obeying the inverse effectiveness rule. Moreover, certain analytic methods can produce results compatible with alternative, or even null, hypotheses.

On the basis on a simulation from the data by Alvarado, Vaughan, Stanford, & Stein (2007), Holmes (2007) claims that if we adopt an absolute measure instead of the relative measure for the MRE (simple difference vs. relative difference in relation to the US baseline), then we would obtain a pattern of results opposite to the IER: a linear relation between in the number of additional spikes in MS conditions as the absolute effectiveness of the best unisensory stimulus.

Similarly, Ross, Saint-Amour, Leavitt, Javitt, & Foxe (2007) assessed evidence for the IER in human audio-visual speech perception, confirming the hypothesis raised by Holmes (2007) and founding evidence of the IER only when the data were expressed as a percentage relative to the US baseline. However, when the gain was expressed as the absolute difference, they found a better performance at the mid-range of stimulus effectiveness. Moreover, when the absolute measure of the multisensory gain was corrected for ceiling effects, the authors reported a sigmoidal increase with increasing stimulus effectiveness.For this reason, here I will provide both absolute and relative measures of the MRE from my experiments, which confirms that the effect I described is a genuine IER (Figure 6.1).

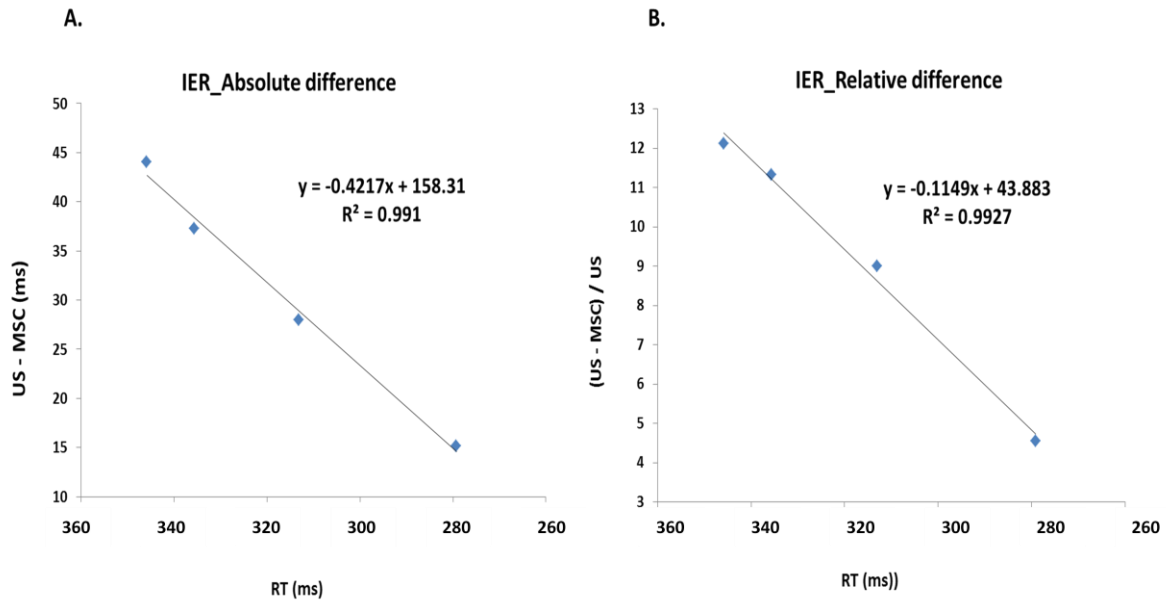


Figure 6.1 Inverse effectiveness rule, Experiment 5. Data from Experiment 5 were re-analysed in order to compare them with the result from other studies in literature. The gain in the AV condition expressed both as absolute (A, simple subtraction of RTs between MS and US condition) or relative (B) difference (subtraction of RTs between AV and V condition, relative to the US baseline. The same inverse relation is evident between

the best US response (abscissa) and the MRE (ordinate) expressed either as an absolute or relative difference.

Figure 6.1 shows that, independently of the absolute vs. relative measure, the pattern of data displays a negative relation between stimulus effectiveness (expressed as speed of US response) and the MRE.

What can be the reason of the difference between my data and those by Holmes (2007) and Ross et al. (2007)? In the first place, it is important to point out the basic difference in the stimuli and approach adopted, that makes difficult any detailed comparison: number of spikes in the case of Holmes (2007), speech perception in the case of Ross et al. (2007), and simple audio-visual stimuli in my experiments. We could assume that any integration that involves speech perception should be fundamentally different from the perception of more simple stimuli, since different brain areas are involved. In fact, the SC is a structure specialized for detection and orientation, a function that intuitively benefits more from inverse effectiveness. The speech recognition, on the contrary, requires higher level semantic elaboration and not just the simple presence or absence of a speech stimulus. Thus, it is plausible to find considerable differences between the processing of the speech and that found in the SC.

Further conclusions are not possible on the basis of present data, and additional experiments are needed to clarify the IER at the behavioural level, and put more light on the different result present in literature.

Post-Hoc Vs. a Priori analysis of the IER

My results regarding the IER are robust toward another criticism highlighted by Holmes (2009). As the reader may have noticed, there is a discrepancy between the definition provided of the IER, the analysis performed in Chapter 3 and the graphs presented in the present paragraph. In fact, the IER is defined as the inverse relationship between the MRE and the *best US response*, as plotted in Figure 6.1. However, the Figure 3.8 in Chapter 3 reports the luminance value in the abscissa instead. In fact, the two analysis represents two alternative methods for the analysis of the IER: the *post-hoc* analysis (related to the US response) and the *a priori* analysis (related to characteristics of

the stimuli; e.g. luminance contrast). With the post-hoc or response-related method, the measure of multisensory integration (y-axis) is regressed on the unisensory response (x-axis) (e.g., Kayser et al. 2008; Longo et al. 2008; Perrault et al. 2003). In the a priori or stimulus-related method, the multisensory response in plotted against several levels of unisensory stimulus intensity, from weak (e.g. low luminance contrast, low saliency) to relatively strong stimuli (high luminance contrast, high saliency; e.g., Alvarado et al. 2007a, b; Longo et al. 2008; Meredith and Stein 1986b; Perrault et al. 2003; Ross et al. 2007; Stevenson and James 2009).

Using this design, there is no need to re-sort the data according to a post-hoc criterion, since different levels of stimulus intensity were built into the experimental design. In his study, Holmes (2009) suggests that the data should not be re-sorted by the obtained responses (e.g., Holmes 2007), in order to avoid the effect of the *regression toward the mean*.

In order to provide strong evidence of the IER in my data, bot analysis were presented, a priori in Chapter 3, and post-hoc (with both relative and absolute measures) in the present paragraph. As a consequence, the three approaches demonstrated the same pattern of data, proving a genuine IER effect.

6.1.4 Multisensory integration in the saccadic system

The starting point for the empirical work took advantage of the rich literature on MI in the generation of saccades. On the other hand, research on saccadic inhibition (Reingold & Stampe, 2002) provided the ground for linking the two paradigms (MI and SI) in order to shed light on MI within the processes of target selection operated by the SC. The purposes of linking the SI to MI are multiple. Firstly, the SI may provide a basis for the study of MI for non-target stimuli (distractors) in the process of target selection, which has been mostly studied for targets so far. Secondly, it allows us to investigate the influence that various distractors may have on the mechanism of MI operating with bimodal targets. Moreover, since the SI is known to occur at very short latencies (60-70 ms), it can help us to delineate the temporal sequence of MI as well. In conclusion, distractor interference can reveal what type of information the saccadic system is sensitive to, at different points in time.

Experiments 7 and 9 from Chapter 6 suffered from a few methodological weaknesses, so Experiments 8 and 10 were run with the aim of improving the experimental design. Overall, the experiments addressed the question of the influence of MI on saccade generation, both when the multisensory stimulus was a distractor (AV distractor, Experiments 7 and 8) and when it was the target of the saccade (AV target and V distractors, Experiments 9 and 10). Experiments 7 and 8 showed that AV distractors are more efficient in inhibiting a saccadic response toward a visual target, in comparison to A or V distractor alone (the difference is expressed both as an exaggeration of the SRT cost, and as an increase in SI dip magnitude). The increase of inhibition for multisensory distractors, however, did not significantly exceed the Race model boundary, suggesting that it can be explained by statistical facilitation and that the two stimuli did not combine to form a more powerful inhibitory stimulus.

In Experiments 9 and 10 the AV stimulus was presented in the target position, with a competing visual distractor on the contralateral side. The assumption was that, since multisensory targets have robust effects on the initiation time of saccades (Arndt & Colonius, 2001; Corneil, Van Wanrooij, Munoz, & Van Opstal, 2002; Engelken & Stevens, 1989) such targets may have similar effects on other aspects of saccadic programing, and make them more resistant to distractors, e.g. by creating a more robust saccadic target representation. However, the results from Experiments 9 and 10 did not support this hypothesis. Instead, both effects (inhibition and MI) seemed to occur independently within the same experimental design. In Experiment 10, saccades toward AV stimuli were faster than those to unisensory stimuli in both conditions in which the distractor was present or absent. However, let me point out that only the MI in the “non- distractor” condition violated the Race model prediction, suggesting that the distractor prevented the occurrence of MI, through lateral inhibitory projections.

On the other hand, the distracting effect was present in auditorily, visually and audio-visually driven saccades in comparison to the three baseline conditions in which no distractor was present (with A, V and AV targets). This pattern of results suggested that the multisensory AV target influenced only the speed of the SRT, while other aspects of the motor programming remained unaffected. In fact, in order to rule out the possibility that the AV targets will be processed faster and will therefore skip the inhibition of the

distractors (since their activations are not overlapping in time), the distractors were "time-locked" to the median time needed to trigger a saccade, so that the onset of the distractor occurred 90 ms after the onset of each distractor.

In the light of the results here summarised, a few theoretical questions arise, and I will try to discuss them in the following paragraphs.

6.2.4.1 SI and Multisensory depression: same mechanism?

An interesting aspect of the data from Chapter 6 regards the inhibition caused by an auditory distractor in Experiments 7 (evident only as an inhibition dip in the distribution) and in 8 (evident both from SRT and SI analysis). To sum up, when the auditory distractor was positioned contralaterally to visual target (Experiment 7, 19.8° interstimulus distance between the visual target and auditory distractor) or on the top and bottom of the screen (Experiment 8), a significant SI was evident.

If we assume that the SI is a multisensory phenomenon, then we can hypothesise that this inhibition may actually represent what is called "multisensory depression" by other authors. Multisensory depression refers to a significant decrease in the neuron's response to the cross-modal stimulus pair as compared with the most effective of these stimuli individually (Kadunce et al. 1997; Meredith and Stein 1985, 1986a,b). It is generally observed when these sensory stimuli are spatially disparate (e.g., derived from different events) such that one is eccentric to its receptive field and the other is within its receptive field. It has been noted in multisensory cortical neurons of animals (Wallace et al. 1992) and in fMRI studies of human cortex (Laurienti et al. 2002). However, this effect has not always been supported by behavioural data in humans. For example, Frens et al. (1995) are among those who systematically varied the spatial discrepancies between the visual target and the auditory accessory stimulus. As already described, they did not obtain any evidence of depression (or inhibition, increase of SRT) even with large interstimulus discrepancies (54°). The same result was obtained by Hughes et al. (1998) with an interstimulus distance of 30°. In light of these results, it seems that the spatial discrepancy alone is not enough to produce a multisensory depression of the behavioural response. However, it is important to note that none of them reached the interstimulus distance of 60° set by Stein et al. (1988) in the experiment investigating the multisensory depression at

neuronal level. Similarly, other studies have manipulated the temporal disparity between the A distractor and V target, also finding no explicit proof of the multisensory depression. For example, Ross and Ross (1981) did not find evidence of inhibition when the auditory distractor appeared 100 or 300 ms after the visual target.

At this point, I would like to comment on some factors that might have favoured the appearance of the distracting effect in my experiments. For example, it is possible that neither the spatial disparity nor the temporal delay alone is sufficient in eliciting an inhibition, but rather the combination of both (as usually present with visual distractors in the SI literature, as well as in my experiments). In Experiment 2 by Frens et al. (1995) both the time delay (auditory accessory stimulus presented 100 ms after the visual target) and the spatial disparity (with coincident AV stimuli and opposite, with 54° of spatial disparity) were manipulated. Similarly to my experiment, the participants had to ignore the auditory distractor and to focus on the visual target. The authors did not report any increase of SRT when the A stimulus appeared 100 ms after the visual and on the opposite side, but importantly, in this case there was no evidence of a decrease of SRT either. To my knowledge, the only study that can be compared to mine (and by which my Experiment 8 was inspired) is the one by Reingold and Stampe (2004) described in Chapter 5. In their study, they investigated the influence of auditory and visual distractors in a reading task, reporting a clear SI dip with the visual transient but no effect of the auditory stimulus. One possible reason for the lack of the effect in Reingold and Stampe (2004) in the auditory distractor condition, as already commented in Chapter 5, might be the use of a pure tone (2000 Hz square-wave tone) instead of a white noise which might have disadvantaged the localization of the distractor.

Nevertheless, a few studies by Colonius and Arndt (2001) and Diederich and Colonius (2007) do support the idea of an inhibition from an auditory distractor. Both the studies report a prolongation of the SRT when the auditory accessory stimulus (i.e. distractor) was contralateral to the visual target (55°) and appearing after the visual target (120 ms of temporal delay). Diederich and Colonius (2007) reported a “late inhibition”, as they call it, when the auditory distractor was contralateral (40° of interstimulus distance) and appearing 50 ms after the visual target. Moreover, this inhibition was stronger when the distractor was bimodal (audio-somatosensory), concordant with the stronger inhibition

with audio-visual distractors in Experiments 7 and 8 of my thesis. Similarly to my experiment, the participants performed the task in a focused attention paradigm (their responses were always toward the visual stimulus).

It is clear that there are lot of contradictions on this point in the current scientific literature. So what can be the factors that makes auditory accessory stimulus act sometimes as a distractor (and thus inhibiting the response toward a target) and sometimes as a warning stimulus (enhancing the response toward a target)?

Two additional factors will be taken in consideration: direction of the spatial attention and the intensity of the auditory distractor. Interestingly, both factors seems to be in common between my study and the study by Diederich and Colonius (2007), and to differ from those used by Ross and Ross (1981), Fren and Opstal (1995), Hughes et al. (1998) or even Reingold and Stampe (2004). In the case of spatial attention, it was oriented toward the target both in Experiment 7 (target always on the right) and in Experiment 8 (target randomly on the right or left of the central part of the screen, distractor always on the top or bottom of the screen). This was not always the case in the studies reported above. For example, in the studies by Ross and Ross (1981), Fren and Opstal (1995) and Hughes et al. (1998) both the visual target and auditory distractor appeared randomly on the left or right, and spatial attention was thus distributed.

In the case of the intensity, it is known to be an important factor governing MI (Meredith & Stein, 1993). As a general rule, the weaker is the response to unisensory stimuli (and thus, the weaker the target is), the stronger is the integration among them, a phenomenon named the inverse effectiveness rule. If two stimuli are already very efficient on their own (i.e. very intensive), it is more likely that they will be treated as two separate events (Meredith & Stein, 1993). In fact, the auditory stimuli might have been, in some studies, not intensive enough to act as distractors: 5 ms clicks in Hughes et al. (1998); 1000 Hz pure tones in Ross and Ross (1981) and in Reingold and Stampe (2004).

Interestingly, both the intensity and spatial configuration of the auditory distracting stimulus in Diederich and Colonius (2007) was similar to the one in my Experiment 8: a white burst coming simultaneously from two loudspeakers positioned horizontally on the left and right of the fixating point.

To sum up, I might speculate that an auditory stimulus can become a distractor in the generation of saccades when a combination of factors is present: spatial disparity

between the auditory distractor and visual target, temporal delay (with the auditory distractor appearing after the visual target), high intensity of the auditory distractor, and spatial attention directed toward the visual target. This conclusion is purely speculative at this point, and additional experiments are needed, with the aim of systematically varying the above mentioned factors. We can only hypothesise the existence of a mechanism that takes advantage of these factors when it comes to decide about which stimulus is the target of the saccade and which one is the (irrelevant) distractor. If this is the case, and If we assume it to be a mechanism operating in a multisensory context, the same should be true independently of the modality of the distractor and of the target. In fact, the same factors seem to apply also in the opposite situation, with auditory target and visual distractor (Experiments 9 and 10): when the saccade is directed toward an auditory target, a visual distractor prolongs the SRT. Even in this case, most of the factors identified above are present: there is a horizontal spatial disparity between the visual distractor and the auditory target (11° in Experiment 9) or vertical disparity (auditory distractor in the central position, visual distractor from top and bottom of the screen); temporal delay (this time the visual distractors appear after the auditory target); auditory spatial attention oriented toward the auditory target (appearing always on the right in Experiment 9, or always in the centre in Experiment 10); intensity of the distractor (large flashing zone in the case of the visual distractor vs. a localized auditory target).

6.2.4.2. Multisensory Integration only in target position?

The results of Experiments 8 and 10 indicate that MI (defined by the violation of the Race model) occurs when AV stimuli are the target, and does not arise when they are the distractor. What is the reason for this discrepancy?

A possible objection is that the time-course of distraction is much shorter than the one of the target-directed response. If the multisensory enhancement appears late in the response, then the lack of MI that I observed could be due to the lack of time for the MI to occur before the distracting event happened. However, Rowland and Stein (2007) reported that the MI occurs as soon as the information coming from different modalities is available

to the neuron. The authors evaluate the temporal properties of multisensory enhancement in the physiological responses of neurons in the SC, showing that multisensory responses were enhanced from their very onset, and that the acceleration of the enhancement was greatest within the first 40 ms.

I will therefore speculate about a second hypothesis, involving multiple factors.

By intuition, the factors commented in the previous paragraph that made the distractor effective as a distractor, may have influenced the occurrence of MI as well. One possible explanation can be related to the role of spatial attention, which was always oriented toward the AV target (either always on the right in Experiment 8, or to the centre in Experiment 10), and never toward the AV distractor.

An alternative explanation is related to the intensity of the targets and distractors. We might think of an explanation based on the *inverse effectiveness rule*, which states (as already described), that the multisensory stimuli are more likely or effectively integrated when the unisensory responses are weak (Meredith & Stein, 1983).

In the case of my stimuli, we can assume that the visual distractor was already very efficient in inhibiting a response, and the integration with an auditory stimulus was therefore not strong or even absent. In that case, the visual and auditory stimulus are treated as two independent distractors, and the inhibition is caused by the faster distractor (according to the Race model of the MI, and to the absence of violation of the Race model; Miller, 1982). On the contrary, the target in Experiment 10 was less intensive in comparison to the distractor in Experiment 8: a small localized square vs. a large flashing zone. The low intensity and the direction of the spatial attention might have favoured the MI, which resulted in a violation of the race model. However, as already described, the violation of the model disappeared when there were a concomitant visual distractor, indicating that even when the MI is favoured and it occurs, it seems to be very susceptible to lateral inhibitory connections caused by the distractor.

To sum up, on the basis of the experiment above described we can hypothesise that the circumstances that favour MI are: absence of temporal and spatial coincidence, low intensity of the stimuli, orientation of the spatial attention and absence of concomitant inhibiting stimuli.

6.2 Attention and Multisensory integration

I will end this general conclusion chapter with some considerations about the influence of attention on MI. Since in neither of my experiments attention has been directly manipulated, the following considerations are purely speculative and may orient future investigations.

In my thesis, the distinction between focused attention and redundant target paradigms is present across experiments (Chapters 2-4). This distinction is an interesting experimental variation and it may provide an important theoretical perspective. In fact, since the stimuli can be chosen to be physically identical (e.g. Experiments 5 and 6, Chapters 3 and 4) in both paradigms, any difference observed in the corresponding responses has to be due to instructions only.

As a general rule, on the basis of my experiments I would conclude that cross-modally oriented attention is not essential for MI, but can promote some effects, that would not emerge without attention or that appear to be stronger with attention.

The considerations that lead to this conclusion are the following:

I. The spatial rule of MI is evident both when attention is focused to one modality (i.e. visual; Experiments 2 and 3), and when attention is distributed cross-modally, in a redundant target paradigm (Leo et al., 2008). However, the overall amount of facilitation seems to be stronger in the second case.

II. The cross-modal shape congruency between visual shapes and sounds appears when the attention is focused only on the visual modality for some pairings (Experiments 4 and 5). However, the cross-modal distribution of attention seems to promote the congruency effect between pairs of stimuli that did not emerge in a focused attention paradigm (Experiment 6).

III. In conclusion, when the spatial attention is not oriented toward a distracting (and thus task-irrelevant) AV stimulus (Experiment 8), the interaction is only a statistical

effect, and does not violate the Race model prediction. On the contrary, the A and V stimulus are integrated among them (i.e. violate the Race model) if they appear to be in the centre of the spatial attention in the role of the target of the saccade (Experiment 10).

However, these speculations should be taked with some caution. In fact, it is difficult to control wheather the attention in the focused attention paradigm is effectively oriented toward one modality or it may spread to other modalities. Therefore, the distinction between the redundant target paradagim and focused attention paradigm should not be taken as a clear-cut.

Nevertheles, apart from these methodological considerations, the role of cross-modal attention has been investigated in a study by Mozolic, Hugenschmidt, Peiffer and Laurienti (2008). The authors were interested in how attention impacts the integration of semantically congruent (blue circle and an auditory verbalization of the word ‘‘blue’’) and incongruent, temporally coincident bimodal stimuli. Similarly to my results, they obtained a multisensory gain in both divided and selective attention conditions. Contrary to my results from Chapter 4, the authors reported a violation of the Race model prediction only in the condition in which the attention was divided cross-modally. However, again I have to point that the authors adopted an approach in the orientation of the attention which is fundamentally different from mine: A visual cue, representing an eye or an ear, indicated to participants to selectively attend to a visual or auditory target. In this way, the participant switched between the conditions of divided and selective (visual or auditory) attention within the same experiment, while the attention was blocked across my experiments (focused attention vs. redundant target paradigm).

Overall, beside the methodological differences, our conclusions seem to agree about the fact that the multisensoriality is evident under both conditions, but appear stronger in the case of divided attention. In fact, this conclusion is supported also by the fMRI data from Fairhall & Macaluso (2009), who claimed that both mechanisms are possible: the pre-attentive MI and improvement of MI by endogenous attention.

The authors presented two speaking lips in the two field controlaterally, and the auditory spoken words (congruent only with one of the two speaking lips) centrally,

creating a setting that would usually result in a 'ventriloquist effect' ; i.e. the participants will misallocate the spoken word toward the matching lips. In order to control the locus of visual attention, they additionally asked participants to perform a visual detection task in one hemifield only. In this way, the matching AV configurations could appear either in the centre of the visual attention, or outside the visual attention. They reported that the activity in MI areas, such as the STS and the SC is enhanced when the stimuli are in the focus of the visual spatial attention, only in the case of AV matching stimuli, proving the facilitating role of the endogenous attention.

However, evidence of a pre-attentive interaction was reported as well: the ventriloquist effect took place even when the matching AV configuration was outside the focus of the endogenous attention and resulted in an exogenously attentional shift toward the opposite side, interfering with the detection task.

To conclude, it seems that both directions of interaction are possible between MI and attention, and the circumstances (or the specific task requirements) determine wheatear the modalities will interact automatically before the attention, or will the attention select the stimuli that will afterward interact.

6.3 Conclusions

My thesis aimed at investigating different aspects of the integration between audio and visual stimuli in manual (Chapters 2-4) and saccadic (Chapter 5) responses. The main results are here summarised. In Chapter 2, the spatial rule influenced the strength of the integration between A and V stimuli only when selectively activating the SC. In Chapter 3, the integration between A and V stimuli was modulated by the intensity of the visual stimuli, according to a linear inverse relation between the unisensory response (directly linked to the effectiveness of the visual stimulus) and the amount of MRE. The strength of the integration was, moreover, affected by featural correspondences, such as between the shape of the visual stimulus and the shape of the paired sound. A strong MRE is evident with congruent vs. incongruent matching. Such a congruency effect is present both in a focused attention paradigm (Experiments 5 and 6; Chapters 3 and 4) and in a redundant

target paradigm (Experiment 6, Chapter 4), and seems to be facilitated (at least for some cross-modal pairings) when attention is distributed cross-modally (redundant target paradigm). In Chapter 5, I have assessed the role of MI both when the AV stimulus acted as a distractor (Experiments 7 and 8) and when it acted as the target of the saccade (Experiments 9 and 10). In the first case, I proved that the distracting cost is stronger with the AV distractor, but this interaction does not violate the Race model prediction. In the second case, a strong violation is evident (Experiment 10) only in the condition without the competing visual distractor.

Bibliography

Alais, D., Newell, F.N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving,* 23(1), 3-38.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Curr Biol, 15*(9), 839-843.

Alvarado, J.C., Stanford, T.R., Vaughan, J.W., & Stein, B.E. (2007a). Cortex mediates multisensory but not unisensory integration in superior colliculus. *J Neurosci, 27*(47), 12775-12786.

Alvarado, J.C., Vaughan, J.W., Stanford, T.R., & Stein, B.E. (2007b). Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *J Neurophysiol, 97*(5), 3193-3205.

Amedi, A., Stern, W.M., Camprodon, J.A., Bermpohl, F., Merabet, L., Rotman, S., Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci, 10*(6), 687-689.

Amlôt, R., Walker, R., Driver, J., & Spence, C. (2003). Multimodal visual somatosensory integration in saccade generation. *Neuropsychologia*, *1479*, 1–15.

Andersen, R.A., Snyder, L.H., Bradley, D.C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu Rev Neurosci, 20,* 303-330.

Anstis, S.M., & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In J.D. Mollon, & L.T. Sharpe (Eds) *Colour vision: Psychophysics and physiology*. London: Academic Press.

Arndt, P.A. & Colonius, H. (2003). Two stages in crossmodal saccadic integration: evidence from a visual-auditory focused attention task*. Experimental Brain Research*, *150*, 417–426 .

Barbur, J.L., Harlow, A.J., & Plant, G.T. (1994). Insights into the different exploits of colour in the visual cortex. *Proc Biol Sci, 258*(1353), 327-334.

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: prevalence and familiality. *Perception, 25*(9), 1073-1079.

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: prevalence and familiality. *Perception, 25*(9), 1073-1079.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci, 7*(11), 1190-1192.

Belardinelli, M.O., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., & Romani, G.L. (2004). Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. *Cogn Process, 5*, 167–174.

Bell, A.H., Corneil, B.D., Meredith, M.A., & Munoz, D.P. (2001). The influence of stimulus properties on multisensory processing in the awake primate superior colliculus. *Can J Exp Psychol, 55*(2), 123-132.

Ben-Artzi, E., & Marks, L.E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept Psychophys, 57*(8), 1151-1162.

Ben-Artzi, E., & Marks, L.E. (1999). Processing linguistic and perceptual dimensions of speech: interactions in speeded classification. *J Exp Psychol Hum Percept Perform, 25(3),* 579-595.

Benedek, G., Mucke, L., Norita, M., Albowitz, B., & Creutzfeldt, O. D. (1988). Anterior ectosylvian visual area (AEV) of the cat: physiological properties. *Prog Brain Res, 75*, 245-255.

Bernstein, I.H. (1970). Can we see and hear at the same time? Some recent studies of intersensory facilitation of reaction time. *Acta Psychol (Amst), 33*, 21-35.

Bernstein, I.H., & Edelstein, B.A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J Exp Psychol, 87*(2), 241-247.

Bernstein, I.H., Eason, T.R., & Schurman, D.L. (1971). Hue-tone sensory interaction: a negative result. *Percept Mot Skills, 33*(3), 1327-1330.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys, 29*(6), 578-584.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept Psychophys, 62*(2), 321-332.

Birch, J., Barbur, J.L., & Harlow, A.J. (1992). New method based on random luminance masking for measuring isochromatic zones using high resolution colour displays. *Ophthalmic Physiol Opt, 12*(2), 133-136.

Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp Brain Res, 160*, 273-282.

Bolognini, N., Rasi, F., & Ladavas, E. (2005). Visual localization of sounds. *Neuropsychologia, 43(11),* 1655-1661.

Bolognini, N., Rasi, F., Coccia, M., & Ladavas, E. (2005). Visual search improvement in hemianopic patients after audio-visual stimulation. *Brain, 128*(12), 2830-2842.

Bompas, A. & Sumner, P. (2009). Temporal dynamics of saccadic distraction. *Journal of Vision 9(9):17*, 1-14.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Curr Biol, 17*(19), 1697-1703.

Born, S., & Kerzel, D. (2008). Influence of target and distractor contrast on the remote distractor effect. *Vision Res, 48*(28), 2805-2816.

Bradley, A., Switkes, E., & De Valois, K. (1988). Orientation and spatial frequency selectivity of adaptation to color and luminance gratings. *Vision Res, 28*(7), 841-856.

Bruce, C.J., Goldberg, M.E., Bushnell, M.C., & Stanton, G.B (1985) Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J Neurophysiol*, *54 (3):*714-34.

Bucur, B., Allen, P.A., Sanders, R.E., Ruthruff, E., & Murphy, M.D. (2005). Redundancy gain and coactivation in bimodal detection: Evidence for the preservation of coactive processing in older adults. *Journals of Gerontology, 60B*, P279-P282.

Buonocore, A., & McIntosh, R.D. (2008). Saccadic inhibition underlies the remote distractor effect. *Experimental Brain Research*, *191 (1*):117-22

Burnett, L.R., Stein, B.E., Chaponis, D., & Wallace, M. T. (2004). Superior colliculus lesions preferentially disrupt multisensory orientation. *Neuroscience, 124*, 535-547.

Burton, H., & Kopf, E.M. (1984). Ipsilateral cortical connections from the second and fourth somatic sensory areas in the cat. *J Comp Neurol, 225*(4), 527-553.

Bushara, K.O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci, 21*(1), 300-304.

Calkins, D.J. (2001). Seeing with S cones. *Prog Retin Eye Res, 20*, 255-287.

Callan, D.E., Callan, A.M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. [Clinical Trial]. *Brain Res Cogn Brain Res, 10*(3), 349-353.

Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*(2), 427-438.

Calvert, G.A., Spence, C., & Stein, B.E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Cassin, B. and Solomon, S. Dictionary of Eye Terminology. Gainsville, Florida: Triad Publishing Company, 1990

Cavanagh, P. (1991). Vision at equiluminance. In J. J. Kulikowski, I. J. Murray & V. Walsh (Eds.), *Vision and Visual Dysfunction: Limits of Vision.* (Vol. 5, pp. 234-250). Boca Raton, FL: CRC Press.

Cavanagh, P., Adelson, E.H., & Heard, P. (1992). Vision with equiluminant colour contrast: 2. A large-scale technique and observations. *Perception, 21*(2), 219-226.

Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *J Exp Psychol Hum Percept Perform, 16*(3), 479-491.

Cavanagh, P., MacLeod, D.I., & Anstis, S.M. (1987). Equiluminance: spatial and temporal factors and the contribution of blue-sensitive cones. *J Opt Soc Am A, 4*(8), 1428-1438.

Chatterjee, S., & Callaway, E.M. (2002). S cone contributions to the magnocellular visual pathway in macaque monkey. *Neuron, 35*(6), 1135-1146.

Clarey, J.C., & Irvine, D.R. (1986). Auditory response properties of neurons in the anterior ectosylvian sulcus of the cat. *Brain Res, 386*(1-2), 12-19.

Clark, H.H., & Brownell, H.H. (1976). Position, direction, and their perceptual integrality. *Perception & Psychophysics, 19*, 328–334.

Clemo, H.R., & Stein, B.E. (1982). Somatosensory cortex: a 'new' somatotopic representation. *Brain Res, 235*(1), 162-168.

Clemo, H.R., & Stein, B.E. (1983). Organization of a fourth somatosensory area of cortex in cat. *J Neurophysiol, 50*(4), 910-925.

Colby, C.L., Duharnel, J.R., & Goldberg, M.E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophvsiology*,*5*, 2841-2852

Colonius, H. & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies . *Perception and Psychophysics*, *63*, 126 – 147 .

Colonius, H. & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16 (6),* 1000 – 1009

Corballis, M.C. (1998). Interhemispheric neural summation in the absence of the corpus callosum. *Brain, 121 (Pt 9)*, 1795-1807.

Corneil, B.D., & Munoz, D.P. (1996). The influence of auditory and visual distractors on human orienting gaze shifts. *J Neurosci, 16(24):*8193-207

Corneil, B.D., Van Wanrooij, M., Munoz, D.P., & Van Opstal, A.J. (2002) Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology, 88*:438–454.

Corney, D., Haynes J.D., Rees G., & Lotto R.B. (2009). The brightness of colour. *PLoS ONE.;4*:e5091.

Coward, S.W., & Stevens, C.J. (2004). Extracting meaning from sound: Nomic mappings, everyday listening, and perceiving object size from frequency. *The Psychological Record, 54 (3)*, 349-364.

Dacey, D.M., & Lee, B.B. (1994). The ‘‘blue on’’ opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature, 367*, 731–735.

de Jong, J.J., Hodiamont, P.P., & de Gelder, B. (2010). Modality-specific attention and multisensory integration of emotions in schizophrenia: reduced regulatory effects. *Schizophr Res, 122*(1-3), 136-143.

de Monasterio, F.M. (1978). Properties of ganglion cells with atypical receptive-field organization in retina of macaques. *J Neurophysiol, 41*(6), 1435-1449.

De Valois, R.L. (1960). Color vision mechanisms in the monkey. *J Gen Physiol, 43(6)Suppl*, 115-128.

Deubel, H., Wolf, W., & Hauske, G., (1984). The evaluation of oculomotor error signals. In G. A. G. & J. F. (Eds.), *Theoretical and Applied Aspects of Oculomotor Research* (pp. 55–62). Amsterdam: Elsevier.

DeYoe, E.A., & Van Essen, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci, 11*(5), 219-226.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? *Psychol Res, 49*, 23–29.

Diederich, A., & Colonius, H. (2007). Why two ‘Distractors’ are better than one: modelling the effect of non-target auditory and tactile stimuli on visual saccadic reaction time. *Experimental Brain Research*, *179 (1),* 43–54.

Diederich, A., & Colonius, H. (2008). Crossmodal interaction in saccadic reaction time: separating multisensory from warning effects in the time window of integration model. *Exp Brain Res, 186*(1), 1-22.

Diederich, A., Colonius, H., Bockhorst, D., & Tabeling, S. (2003) Visual-tactile spatial interaction in saccade generation. *Experimental Brain Research*, *148(3):*328-37

Dobkins, K.R., & Albright, T.D. (1994). What happens if it changes color when it moves?: the nature of chromatic input to macaque visual area MT. *J Neurosci, 14*(8), 4854-4870.

Dodd, B. (1977). The role of vision in the perception of speech. *Perception, 6*(1), 31-40.

Dowling, W.J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), The psychology of music. New York: Academic Press.

Driver, J. (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381*, 66–68.

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philos Trans R Soc Lond B Biol Sci, 353*(1373), 1319-1331.

Drum, B. (1983). Short-wavelength cones contribute to achromatic sensitivity. *Vision Res, 23*(12), 1433-1439.

Edelman, J.A., and Xu, K.Z. (2009) Inhibition of voluntary saccadic eye movement commands by abrupt visual onsets. *J Neurophysiol, 101,* 1222–1234.

Eisner, A., & MacLeod, D.I. (1980). Blue-sensitive cones do not contribute to luminance. *J Opt Soc Am, 70*(1), 121-123.

Engelken, E.J., & Stevens, K.W. (1989). Saccadic eye movements in response to visual, auditory, and bisensory stimuli. *Aviat Space Environ Med*, *60(8):*762-8.

Ernst, M.O., & Bulthoff, H.H. (2004). Merging the senses into a robust percept. *Trends Cogn Sci, 8*(4), 162-169.

Evans, K.K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J Vis, 10*(1).

Fairhall, S.L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur J Neurosci, 29(6),* 1247-1257.

Favreau, O.E., & Cavanagh, P. (1981). Color and luminance: independent frequency shifts. *Science, 212*(4496), 831-832.

Felleman, D.J., & Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex, 1(1), 1-47.

Findlay, J.M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioural Brain Science*, *22*:661–675.

Fishbach, A., Nelken, I., and Yeshurun, Y. (2001) Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. J. Neurophysiol. 85, 2303-2323.

Flanagan, P., Cavanagh, P., & Favreau, O.E. (1990). Independent orientation-selective mechanisms for the cardinal directions of colour space. *Vision Res, 30*(5), 769-778.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., & Schroeder, C.E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res, 10*(1-2), 77-83.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res, 147*, 332-343.

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Ladavas, E. (2005). Audiovisual integration in patients with visual deficit. *J Cogn Neurosci, 17*(9), 1442-1452.

Frens, M.A., & Van Opstal, A.J. (1998) Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Research Bulletin*, *46*:211–224.

Frens, M.A., Van Opstal, A.J., & Van der Willigen, R.F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception and Psychophysics, 57*, 802 – 816.

Frisoli, A., & Camurri, A. (2006). Special Issue Editorial: Multisensory interaction in virtual environments. *Virtual Reality, 10*(1), 2-3.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Percept Psychophys, 68*(7), 1191-1203.

Gallant, J.L., Braun, J., and Essen, D.C.V. (1993). Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex.Science, 259:100–103.

Garner, W.R. (1974). *The processing of information and structure.* New York: Wiley.

Garner, W.R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*, 98–123.

Garrido, M.I, Kilner, J.M., Stephan, K.E., and Friston, K.J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clin Neurophysiol, 120(3):* 453–463.

Gaver, W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*, 1-29.

Gegenfurtner, K.R., Kiper, D.C., Beusmans, J.M., Carandini, M., Zaidi, Q., & Movshon, J.A. (1994). Chromatic properties of neurons in macaque MT. *Vis Neurosci, 11*(3), 455-466.

Giard, M.H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci, 11*(5), 473-490.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gielen, S.C., Schmidt, R.A., & Van den Heuvel, P.J. (1983). On the nature of intersensory facilitation of reaction time. *Percept Psychophys, 34*(2), 161-168.

Ginsberg, L. (1923). A case of synaesthesia. *Am. J. Psychol., 34*, 582–589.

Godijn, R., & Theeuwes, J. (2002). Oculomotor capture and Inhibition of Return: evidence for an oculomotor suppression account of IOR. [Clinical Trial]. *Psychol Res, 66*(4), 234-246.

Gottlieb, J.P., Kusunoki, M., & Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 39 1(6666*), 48 1-484.

Gouras, P. (1968). Identification of cone mechanisms in monkey ganglion cells. *J Physiol, 199.*

Graupner, S.T., Velichkovsky, B.M., Pannasch, S., & Marx, J. (2007) Surprise, surprise: two distinct components in the visually evoked distractor effect. *Psychophysiology, 44(2):*251-61.

Gregory, R.L. (1977). Vision with isoluminant colour contrast: 1.A projection technique and observations. *Perception, 6*(1), 113-119.

Groh, J.M., & Sparks, D.L. (1996). Saccades to somatosensory targets. II. motor convergence in primate superior colliculus. *J Neurophysiol, 75*(1), 428-438.

Groh, J.M., & Sparks, D.L. (1996). Saccades to somatosensory targets. I.behavioural characteristics*. Journal of Neurophysiology*,*75(1):*412-27.

Grossenbacher, P.G., & Lovelace, C.T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends Cogn Sci, 5*(1), 36-41.

Hairston, W.D., Laurienti, P.J., Mishra, G., Burdette, J.H., & Wallace, M.T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Exp Brain Res, 152(3),* 404-408.

Hanes, D.P. & Wurtz, R.H. (2001). Interaction of the frontal eye field and superior colliculus for saccade generation. *J. Neurophysiol. 85*: 804-815

Harrington, L.K., & Peck, C.K. (1996) Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*,*122(2*):247-52.

Hendry, S.H., & Reid, R.C. (2000). The koniocellular pathway in primate vision. *Annu Rev Neurosci, 23*, 127-153.

Hepp, K., & Henn, V. (1985). Iso-frequency curves of oculomotor neurons in the rhesus monkey. *Vision Research, 25*,493-499.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *J Exp Psychol, 63*, 289-293.

Hikosaka, O. & Wurtz, R.H. (1983). Visual and oculomotor functions of monkey substantia nigra pars reticulata. II. Visual responses related to fixation of gaze. *J. Neurophysiol. 49*: 1254-1267.

Ho, C., Tan, H.Z. & Spence, C. (2008). Using Spatial Vibrotactile Cues to Direct Visual Attention in Driving Scenes. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*, 397-412.

Hofman, P.M. & Van Opstal, A.J. (1998). Spectro-temporal factors in two-dimensional human sound localization. *Journal of the Acoustical Society of America*, 103, 2634-2648.

Hoggan, E. & Brewster, S. (2007). Designing Audio and Tactile Crossmodal Icons for Mobile Devices. *Proc. Ninth Int'l Conf. Multimodal Interfaces (ICMI '07),* 162-169.

Holland, P.C., & Gallagher, M. (2006). Different roles for amygdala central nucleus & substantia innominata in the surprise-induced enhancement of learning. *Journal of Neuroscience*.5; 26(14):3791-7.Erratum in: Journal of Neuroscience, 26; 26 (17):4715.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. *Neuropsychologia, 45*(14), 3340-3345.

Holmes, N.P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topogr, 21(3-4),* 168-176.

Holmes, N.P., & Spence, C. (2004). The body schema and the multisensory representation(s) of peripersonal space. *Cogn Process, 5*(2), 94-105.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Curr Biol, 15*(18), R762-764.

Horridge, A. (2000). Seven experiments on pattern vision of the honeybee, with a model. Vision Research, 40:2589–2603.

Howard, I.P., & Templeton, W.B. (1966). *Human spatial orientation*. New York: Wiley.

Hsu, Y.F. (2005). A generalization of Piéron's law to include background intensity and latency distribution. *Journal of Mathematical Psychology*, 49(6), 450-463.

Hubbard, E.M., & Ramachandran, V.S. (2005). Neurocognitive mechanisms of synesthesia. *Neuron, 48*(3), 509-520.

Huerta, M.F., and Harting, J.K. (1984). The mammalian superior colliculus: studies of its morphology and connections. In: Vanegas H (ed). Comparative neurology of the optic tectum. Plenum Press, New York, pp 687–773.

Hughes, H.C., Nelson, M.D., & Aronchick, D.M. (1998) Spatial characteristics of visual-auditory summation in human saccades. *Vision Research*, *38*:3955–3963.

Hughes, H.C., Reuter-Lorenz, P.A., Nozawa, G., & Fendrich, R. (1994) Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*. *20(1):*131-53.

Humphreys, G. W., Romani, C., Olson, A., Riddoch, M. J., and Duncan, J. (1994). Non-spatial extinction following lesions of the parietal lobe in humans. Nature, 372:357–359.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *J Autism Dev Disord, 36*(1), 77-90.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

James, T.W., Culham, J., Humphrey, G.K., Milner, A.D., & Goodale, M.A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain, 126*(Pt 11), 2463-2475.

James, T.W., Humphrey, G.K., Gati, J.S., Servos, P., Menon, R.S., & Goodale, M.A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia, 40*(10), 1706-1714.

James, T.W., Kim, S., & Fisher, J.S. (2007). The neural basis of haptic object processing. *Can J Exp Psychol, 61*(3), 219-229.

James, T.W., Stevenson, R.A., Kim, S., Vanderklok, R.M., & James, K.H. (2011). Shape from sound: evidence for a shape operator in the lateral occipital cortex. *Neuropsychologia, 49*(7), 1807-1815.

Jay, M.F., & Sparks, D.L. (1987). Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *J Neurophysiol, 57*(1), 35-55.

Jiang, W., Jiang, H., & Stein, B.E. (2002). Two corticotectal areas facilitate multisensory orientation behavior. *J Cogn Neurosci, 14*(8), 1240-1255.

Jiang, W., Wallace, M.T., Jiang, H., Vaughan, J.W., & Stein, B.E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *J Neurophysiol, 85*(2), 506-522.

Kadunce, D.C., Vaughan, J.W., Wallace, M.T., Benedek, G., & Stein, B.E. (1997). Mechanisms of within- and cross-modality suppression in the superior colliculus. *J Neurophysiol, 78*(6), 2834-2847.

Kaiser, P.K. (1971). Minimally distinct border as a preferred psychophysical criterion in visual heterochromatic photometry. *J Opt Soc Am, 61*(7), 966-971.

Kalesnykas, R.P., & Hallett, P.E. (1994) Retinal eccentricity and the latency of eye saccades. *Vision Research*. *34(4):*517-531.

Kanaseki, T., & Sprague, J.M. (1974). Anatomical organization of pretectal nuclei and tectal laminae in the cat. *J Comp Neurol, 158*, 319-337.

Kaplan, E., & Shapley, R.M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proc Natl Acad Sci U S A, 83*(8), 2755-2757.

Kaplan, E.H., & Lee, Y.S. (1990). How bad can it get? Bounding worst case endemic heterogeneous mixing models of HIV/AIDS. *Math Biosci, 99*(2), 157-180.

Kayser, C., C.I. Petkov, and N.K. Logothetis, (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex, 18*:1560–74.

King, A.J., & Hutchings, M.E. (1987). Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: a map of auditory space. *J Neurophysiol, 57*(2), 596-624.

King, A.J., & Palmer, A.R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp Brain Res, 60*, 492-500.

King, W.M., Lisberger, S.G., & Fuchs, A.F. (1986). Oblique saccadic eye movements of primates. *Journal of Neurophvsiologv*, *56*, 769-784.

King-Smith, P.E., & Carden, D. (1976). Luminance and opponent-color contributions to visual detection and adaptation and to temporal and spatial integration. *J Opt Soc Am, 66*(7).

Köhler, W. (1929). *Gestalt Psychology*. New York: Liveright.

Kopecz, K. (1995). Saccadic reaction times in gap/overlap paradigms: a model based on integration of intentional and visual information on neural, dynamic fields. *Vision Res, 35*(20), 2911-2925.

Kraus, N., McGee, T., Littman, T., & Nicol, T. (1992) Reticular formation influences on primary and non-primary auditory pathways as reflected by the middle latency response. *Brain Research*. *7;587*(2):186-94.

Krummenacher, J., Müller, H. J., & Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel coactive processing of dimensions. *Perception & Psychophysics, 63*, 901-917.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., & Schroeder, C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53(2),* 279-292.

Langendijk, E. & Bronkhorst, A.W. (2002). Contribution of spectral cues to human sound localization. *Journal of the Acoustical Society of America*, *112*, 1583-1596.

Langendijk, E., Kistler, D.,J., & Wightman, F.L. (2001). Sound localization in the presence of one or two distractors. *Journal of the Acoustical Society of America*, 109, 2123-2134.

Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., & Wallace, M.T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res, 158*(4), 405-414.

Lawson, K.R., & Turkewitz, G. (1980). Intersensory function in newborns: effect of sound on visual preferences. *Child Dev, 51*(4), 1295-1298.

Lee, C., Chung, S., Kim, J., and Park, J. (1991). Auditory facilitation of visually guided saccades. *Society of Neuroscience Abstract*, 17, 862.

Lee, J., & Stromeyer, C.F., (1989). Contribution of human short-wave cones to luminance and motion detection. *J Physiol, 413*, 563-593.

Leo, F., Bertini, C., di Pellegrino, G., & Ladavas, E. (2008). Multisensory integration for orienting responses in humans requires the activation of the superior colliculus. *Exp Brain Res, 186*(1), 67-77.

Lévy-Schoen, A. (1969). Détermination et latence de la réponse oculomotrice à deux stimulus simultanés ou successifs selon leur excentricité relative. *L'Année Psychol, 69*, 373–392.

Lewis, J.W., Wightman, F.L., Brefczynski, J.A., Phinney, R.E., Binder, J.R., & DeYoe, E.A. (2004). Human brain regions involved in recognizing environmental sounds. *Cereb Cortex, 14*(9), 1008-1021.

Lewald, J., Dörrscheidt, G.J., & Ehrenstein, W.H. (2000). Sound localization with eccentric head position. *Behavioural Brain Research, 108*. 105–125.

Li, J.C., Sampson, G.P., & Vidyasagar, T.R. (2007). Interactions between luminance and colour channels in visual search and their relationship to parallel neural channels in vision. *Exp Brain Res, 176*(3), 510-518.

Livingstone, M.S., & Hubel, D.H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J Neurosci, 7*(11), 3416-3468.

Lockhead, G.R. (1972). Processing dimensional stimuli: a note. *Psychol Rev, 79*(5), 410-419.

Logothetis, N.K., & Charles, E.R. (1990). The minimum motion technique applied to determine isoluminance in psychophysical experiments with monkeys. *Vision Res, 30*(6), 829-838.

Long, J. (1976). Divided attention, order of report and responses to immediately successive signals. *Q J Exp Psychol, 28(3)*, 339-353.

Longo, M. R., Cardozo, S., & Haggard, P. (2008). Visual enhancement of touch and the bodily self. *Conscious Cogn, 17*(4), 1181-1191.

Ludwig, C.J., Gilchrist, I.D., & McSorley, E. (2004). The influence of spatial frequency and contrast on saccade latencies. *Vision Res, 44*(22), 2597-2604.

Ludwig, C.J., Gilchrist, I.D., & McSorley, E. (2005). The remote distractor effect in saccade programming: channel interactions and lateral inhibition. *Vision Res, 45*(9), 1177-1190.

Lueck, C.J., Crawford, T.J., Savage, C.J., & Kennard, C. (1990). Auditory-visual interaction in the generation of saccades in man. *Experimental Brain Research*, *82*, 149–157.

Lynch, J.C., Graybiel, A.M., & Lobeck, L.J. (1985). The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. *Journal of Comparative Neurology*, *235*(2), 24 1-254.

Macaluso, E., Frith, C.D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289*(5482), 1206-1208.

Maravita, A., Bolognini, N., Bricolo, E., Marzi, C.A., & Savazzi, S. (2008). Is audiovisual integration subserved by the superior colliculus in humans? *Neuroreport, 19*, 271-275.

Marks, L.E. (1978). Binaural summation of the loudness of pure tones. *J Acoust Soc Am, 64*(1), 107-113.

Marks, L.E. (1982a). Bright sneezes and dark coughs, loud sunlight and soft moonlight. *J Exp Psychol Hum Percept Perform, 8*(2), 177-193.

Marks, L.E. (1982b). Synesthetic perception and poetic metaphor. *J Exp Psychol Hum Percept Perform, 8*(1), 15-23.

Marks, L.E. (1987). On cross-modal similarity: auditory-visual interactions in speeded discrimination. *J Exp Psychol Hum Percept Perform, 13*(3), 384-394.

Marks, L.E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *Int J Psychophysiol, 50*(1-2), 125-145.

Marks, L.E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monogr Soc Res Child Dev, 52*(1), 1-102.

Marrocco, R.T., & Li, R.H. (1977). Monkey superior colliculus: properties of single cells and their afferent inputs. *J Neurophysiol, 40*(4), 844-860.

Martino, G., & Marks, L.E. (1999). Perceptual and linguistic interactions in speeded classification: tests of the semantic coding hypothesis. *Perception, 28*(7), 903-923.

Martino, G., & Marks, L.E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence. *Perception, 29*(6), 745-754.

Marzi, C.A., Smania, N., Martini, M.C., Gambina, G., Tomelleri, G., Palamara, A., & Prior, M. (1996). Implicit redundant-targets effect in visual extinction. *Neuropsychologia, 34*(1), 9-22.

Maurer, D. (1997). Neonatal synaesthesia: implications for the processing of speech and faces. In S. Baron-Cohen & J. E. Harrison (Eds.), *Synaesthesia: Classic and Contemporary Readings* (pp. 224–242). Oxford: S. Blackwell.

McDonald, J.J., Teder-Salejarvi, W.A., & Hillyard, S.A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407* (6806), 906-908.

McDonald, J.J., Teder-Salejarvi, W.A., Di Russo, F., & Hillyard, S.A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *J Cogn Neurosci, 15*(1), 10-19.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

McHaffie J.G., Kruger L., Clemo H.R. & Stein B.E. (1988). Corticothalamic and corticotectal somatosensoryprojections from the anterior ectosylvian sulcus (SIV cortex) inneonatal cats: an anatomical demonstration with HRP and 3H-leucine. *J Comp Neurol. 274(1):*115-126.

McIlwain, J.T. (1988). Effects of eye position on electrically evoked saccades: a theoretical note. *Vis Neurosci*, *1*(2):239-44.

McKenna, T.M., Diamond, D.M., & Weinberger, N. M. (1984). Response patterns ot single audi-tory cortical neurons to tone sequences. *Society for Neuroscience Abstracts, 10*, 244.

Melara, R.D. (1989). Dimensional interaction between color and pitch. *J Exp Psychol Hum Percept Perform, 15*(1), 69-79.

Melara, R.D., & Mounts, J.R.W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics, 56*, 73-90.

Melara, R.D., & O'Brien, T.P. (1987). Interaction between synesthetically corresponding dimensions. *J Exp Psych General, 116*, 323–336.

Meltzoff, A.N., & Borton, R.W. (1979). Intermodal matching by human neonates. *Nature, 282(5737)*, 403-404.

Meredith, M.A., & Clemo, H.R. (1989). Auditory cortical projection from the anterior ectosylvian sulcus (Field AES) to the superior colliculus in the cat: an anatomical and electrophysiological study. *J Comp Neurol, 289*(4), 687-707.

Meredith, M.A., & Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389-391.

Meredith, M.A., & Stein, B.E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res, 365*(2), 350-354.

Meredith, M.A., & Stein, B.E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol, 56*, 640-662.

Meredith, M.A., & Stein, B.E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol, 75*(5), 1843-1857.

Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci., 7*, 3215–3229

Merigan, W.H., & Maunsell, J.H. (1993). How parallel are the primate visual pathways? *Annu Rev Neurosci, 16*, 369-402.

Michelson, A. (1927). *Studies in Optics*. Chicago: University of Chicago Press.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol, 14*(2), 247-279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Percept Psychophys, 40*(5), 331-343.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *J Exp Psychol Hum Percept Perform, 17*(1), 160-169.

Mollon, J.D. (1982). Color vision. *Annu Rev Psychol, 33*, 41-85.

Mondloch, C.J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cogn Affect Behav Neurosci, 4*(2), 133-136.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Brain Res Cogn Brain Res, 17*(1), 154-163.

Morrell, F. (1972). Visual system's view of acoustic space. *Nature, 238*(5358), 44-46.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Mozolic, J.L., Joyner, D., Hugenschmidt, C.E., Peiffer, A.M., Kraft, R.A., Maldjian, J.A., & Laurienti, P.J. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC Neurol, 8*, 35.

Mullen, K.T. (1991). Colour vision as a post-receptoral specialization of the central visual field. *Vision Res, 31*(1), 119-130.

Munoz, D.P., & Fecteau, J.H. (2002). Vying for dominance: dynamic interactions control visual fixation and saccadic initiation in the superior colliculus. *Prog Brain Res, 140*, 3-19.

Munoz, D.P., & Istvan, P.J. (1998). Lateral inhibitory interactions in the intermediate layers of the monkey superior colliculus. *J Neurophysiol, 79*(3), 1193-1209.

Munoz, D.P., & Wurtz, R.H. (1993). Fixation cells in monkey superior colliculus. II. Reversible activation and deactivation. *Journal of Neurophvsiology* l. *70*(2):576-589.

Munoz, D.P., Waitzman, D.M., & Wurtz, R.H. (1996). Activity of neurons in monkey superior colliculus during interrupted saccades. *Journal of Neurophvsiology*, *75*(6):2562-2580.

Näätänen, R., Teder, W., Alho, K., & Lavikainen, J. (1992). Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport, 3(6*), 493-496.

Navarra, J., Alsius, A., Velasco, I., Soto-Faraco, S., & Spence, C. (2010). Perception of audiovisual speech synchrony for native and non-native language. *Brain Res, 1323*, 84-93.

Nayatani, Y. (1998). A colorimetric explanation of the Helmholtz-Kohlrausch effect. *Color Res Appl. 23*:374–378.

Nickerson, R.S. (1973). Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychol Rev, 80*(6), 489-509.

Oberman, L.M., & Ramachandran, V.S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: the mirror neuron hypothesis. *Soc Neurosci, 3*(3-4), 348-355.

O'Donell, B.M, & Colombo, E.M. (2006). Reaction times to chromaticity, luminance contrast, size, and adaptation luminance changes. [ECVP Abstract Supplement]. *Perception, 35*.

Olivier, E., Dorris M.C., & Munoz D.P. (1999). Lateral interactions in the superior colliculus, not an extended fixation zone, can account for the remote distractor effect. *Behavioral and Brain Sciences*, *22*, 694–695.

Olson, C.R., & Graybiel, A.M. (1987). Ectosylvian visual area of the cat: location, retinotopic organization, and connections. *J Comp Neurol, 261*(2), 277-294.

Ottes, F.P., Van Gisbergen, J.A., & Eggermont, J.J. (1987). Collicular involvement in a saccadic colour discrimination task. *Exp Brain Res, 66*, 465-478.

Pannasch, S., Dornhoefer, S.M., Unema, P.J., & Velichkovsky, B.M. (2001). The omnipresent prolongation of visual fixations: saccades are inhibited by changes in situation and in subject's activity. *Vision Res, 41*(25-26), 3345-3351.

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neurosci Lett, 442*, 257-261.

Parise, C.V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Exp Brain Res, 214*(3), 373-380.

Parise, C.V., & Spence, C. (2008). Synesthetic associations modulate the temporal ventriloquism effect. *Neuroscience Letters, 442*, 257-261.

Parise, C.V., & Spence, C. (2009). 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One, 4*(5).

Patching, G.R., & Quinlan, P.T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *J Exp Psychol Hum Percept Perform, 28*(4), 755-775.

Patt, Y.Z., Hassan, M.M., Lozano, R.D., Waugh, K.A., Hoque, A.M., Frome, A.I., & Raijman, I. (2001). *Clin Cancer Res, 7*(11), 3375-3380.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., and Griffiths, T.D. (2002). The procesing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.

Peck, C.K. (1987). Saccade-related burst neurons in cat superior colliculus. *Brain Res, 408*(1-2), 329-333.

Perrault, T.J., Jr., Vaughan, J.W., Stein, B.E., & Wallace, M.T. (2003). Neuron-specific response characteristics predict the magnitude of multisensory integration. *J Neurophysiol, 90(6),* 4022-4026.

Perron, C., & Hallett, P.E. (1995). Saccades to large coloured targets stepping in open fields. *Vision Res, 35*(2), 263-274.

Platt, P.L., & Glimcher, P.W. (1997). Responses of intraparietal neurons to saccadic targets and visual distractors. *Journal of Neurophvsiology*, *3*, 1574-1589.

Pluta, S.R., Rowland, B.A., Stanford, T. R., & Stein, B. E. (2011). Alterations to multisensory and unisensory integration by stimulus competition. *J Neurophysiol, 106*(6), 3091-3101.

Pomerantz, J.R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*, 516–540.

Pridmore, R.W. (2007). Effects of luminance, wavelength and purity on the color attributes: Brief review with new data and perspectives. *Color Research & Application, 32*:208–222.

Raab, D.H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24*, 574-590.

Rafal, R., Smith, J., Krantz, J., Cohen, A., & Brennan, C. (1990) Extrageniculate vision in hemianopic humans: saccade inhibition by signals in the blind field. *Science*, *5*;250(4977):118-21.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron, 28*(2), 617-625.

Ramachandran, V.S., & Hubbard, E.M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proc Biol Sci, 268*(1470), 979-983.

Ramachandran, V.S., & Hubbard, E.M. (2003). Hearing colors, tasting shapes. *Sci Am, 288*(5), 52-59.

Rayner, K., Castelhano, M.S., & Yang, J. (2009). Eye movements and the perceptual span in older & younger readers. *Psychol Aging*. *24(3)*:755-760.

Reingold, E.M., & Stampe, D.M. (1999). Saccadic inhibition in complex visual tasks. In W. Becker, H. Deubel & T. Mergner (Eds.), *Current oculomotor research: physiological and psychological aspects* (pp. 249–255). London: Plenum.

Reingold, E.M., & Stampe, D.M. (2000) Saccadic inhibition and gaze contingent research paradigms. In: A. Kennedy, R. Radach, D. Heller, J. Pynte (eds) Reading as a perceptual process. Elsevier, Amsterdam, 119–145.

Reingold, E.M., & Stampe, D.M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *J Cogn Neurosci,* 371-388.

Reingold, E.M., & Stampe, D.M. (2003). Using the saccadic inhibition paradigm to investigate saccadic control in reading. In: Hyona J, Radach R, Deubel H (eds). The mind's eye: cognitive and applied aspects of eye movements. Elsevier, Amsterdam, pp 347–360.

Reingold, E.M., & Stampe, D.M. (2004) Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance*. *30*:194–211.

Reisberg, D., & Leak, S. (1987). Visual imagery and memory for appearance: Does Clark Gable or George C. Scott have bushier eyebrows? *Canadian Journal of Psychology, 41*, 521-526.

Reuter-Lorenz, P.A., Nozawa, G., Gazzaniga, M.S., & Hughes, H.C. (1995). Fate of neglected targets: a chronometric analysis of redundant target effects in the bisected brain. *J Exp Psychol Hum Percept Perform, 21*(2), 211-230.

Robinson, D.A. (1970). Oculomotor unit behavior in the monkey. *Journal of Neurophvsiology, 3*,393-404.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C., & Foxe, J.J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr Res, 97*(1-3), 173-183.

Ross, S.M., & Ross, L.E. (1981) Saccade latency and warning signals: effects of term and short-term monocular deprivation upon oculomotor function in the auditory and visual offset and onset. *Perception and Psychophysics*, *29*:429–437.

Ross, L. E., & Ross, S. M. (1980). Saccade latency and warning signals: stimulus onset, offset, and change as warning events. *Perception & Psychophysics, 27*,251-257.

Russo, G.S. & Bruce, C.J. (1994). Frontal eye field activity preceding aurally guided saccades. *Journal of Neurophysiology*, *71*(3):1250-1253.

Rusconi, E., Kwan, B., Giordano, B.L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition, 99 (2),* 113–129.

Sagiv, N. (2004). Synesthesia in perspective. In L. C. Robertson & N. Sagiv (Eds.), *Synesthesia: Perspectives from Cognitive Neuroscience* (pp. 3–10). New York: Oxford University Press.

Saito, D.N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. (2005). Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. *Cereb Cortex, 15*(11), 1750-1760.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225-239.

Sarko, D.K., Nidiffer, A.R., Powers, III A.R., Ghose, D., Hillock-Dunn, A., Fister M.C., Krueger, J. and Wallace, M.T. (2012). Spatial and Temporal Features of Multisensory Processes: Bridging Animal and Human Studies. In: Murray MM, Wallace MT, (eds). The Neural Bases of Multisensory Processes. Boca Raton (FL): CRC Press.

Savazzi, S., & Marzi, C.A. (2002). Speeding up reaction time with invisible stimuli. *Curr Biol, 12*(5), 403-407.

Savazzi, S., & Marzi, C.A. (2004). The superior colliculus subserves interhemispheric neural summation in both normals and patients with a total section or agenesis of the corpus callosum. *Neuropsychologia, 42*(12), 1608-1618.

Schiller, P.H., & Malpeli, J.G. (1977). Functional specificity of lateral geniculate laminae in the rhesus monkey. J Neurophysiol, 40, 428-445.

Schröger, E. (1997). On the detection of auditory deviations: a pre-attentive activation model.. *Psychophysiology, 34(3),* 245-257.

Schwarz, W., & Ischebeck, A. (1994). Coactivation and statistical facilitation in the detection of lines. *Perception, 23*(2), 157-168.

Segraves, M.A., & Goldberg, M.E. (1987). Functional properties of corticotectal neurons in the monkey's frontal eye field. *Journal of Neurophysiology*, *58*(6), 1387-1419.

Serino, A., Farne, A., Rinaldesi, M. L., Haggard, P., & Ladavas, E. (2007). Can vision of the body ameliorate impaired somatosensory function? *Neuropsychologia, 45(5)*, 1101-1107.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408*(6814), 788.

Shapley, R. (1990). Visual sensitivity and parallel retinocortical channels. *Annu Rev Psychol, 41*, 635-658.

Shiu, L., & Pashler, H. (1994). Negligible effect of spatial precueing on identification of single digits. *J Exp Psychol Hum Percept Perform, 20*, 1037–1054.

Shook, B.L., Schlag-Rey, M., & Schlag, J. (1990). Primate supplementary eye field: I.Comparative aspects of mesencephalic and pontine connections. *Journal of Comparative Neurology*, *301*(4), 618-642.

Shore, D.I., & Simic, N. (2005). Integration of visual and tactile stimuli: top-down influences require time. *Exp Brain Res, 166*(3-4), 509-517.

Small, D.M., Voss, J., Mak, Y.E., Simmons, K.B., Parrish, T., & Gitelman, D. (2004). Experience-dependent neural integration of taste and smell in the human brain. *J Neurophysiol, 92*, 1892-1903.

Smithson, H. (2005). What's special about s-cone vision? *Journal of Vision*, 2005, 5(12) (abstract).

Soto-Faraco, S., Sinnett, S., Alsius, A., & Kingstone, A. (2005). Spatial orienting of tactile attention induced by social cues. *Psychon Bull Rev, 12(6),* 1024-1031.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture. *Cogn Affect Behav Neurosci, 4*(2), 208-217.

Sparks, D.L., & Hartwich-Young, R. (1989). The deep layers of the superior colliculus. *Rev Oculomot Res, 3*, 213-255.

Sparks, D.L., & Hartwich-Young, R. (1989). The deep layers of the superior colliculus. In R.H. Wurtz and M.E. Goldberg (Eds.), The Neurobiology of Saccadic Eve Movements, (pp.213-256).

Sparks, D.L., Holland, R., & Guthrie, B.L. (1976). Size and distribution of movement fields in the monkey superior colliculus. *Brain Res*. *20*;113(1):21-34.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten Percept Psychophys, 73*(4), 971-995.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Percept Psychophys, 59*(1), 1-22.

Spence, C., & Driver, J. (2004). *Cross-Modal Space and Cross-Modal attention*. Oxford: Oxford University Press.

Spence, C., & Walton, M. (2005). On the inability to ignore touch when responding to vision in the crossmodal congruency task. *Acta Psychol (Amst), 118*(1-2), 47-70.

Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cogn Affect Behav Neurosci, 4*(2), 148-169.

Sperling, G., & Dosher, B. A. (1986). In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and performance* (Vol. 1, pp. 2-65). New York: Wiley.

Sperling, H.G., & Harwerth, R.S. (1971). Red-green cone interactions in the increment-threshold spectral sensitivity of primates. *Science, 172*(3979), 180-184.

Sprague, J. M. (1996). Neural mechanisms of visual orienting responses. *Prog Brain Res, 112*, 1-15.

Stanford, T.R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurophysiol*(25), 6499-6508.

Stein, B.E. (1988). Superior colliculus-mediated visual behaviors in cat and the concept of two corticotectal systems. *Prog Brain Res, 75*, 37-53.

Stein, B.E., & Meredith, M.A. (1993). The Merging of the Senses. Cambridge, MA: MIT Press.

Stein, B.E., & Wallace, M.T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Prog Brain Res, 112*, 289-299.

Stein, B.E., Huneycutt, W.S., & Meredith, M.A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Res, 448*(2), 355-358.

Stein, B.E., Meredith, M.A., & Wallace, M.T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog Brain Res, 95*, 79-90.

Stein, B.E., Meredith, M.A., & Wallace, M.T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog Brain Res,*; *95*:79-90.

Stein, B.E., Meredith, M.A., Hunneycutt, W.S., & McDade, L. (1989). Behavioral indices of multisensory integration: orientation of visual cues is affected by auditory stimuli. *J. Cogn. Neurosci., 1*, 12–24.

Stein, B.E., Scott Huneycutt, W., & Alex Meredith, M. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Res, 448(2),* 355-358.

Stein, B.E., Stanford, T.R., Ramachandran, R., Perrault, T.J., Jr., & Rowland, B.A. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Exp Brain Res, 198*(2-3), 113-126.

Stevens, J.C., & Marks, L.E. (1965). Cross-modality matching of brightness and loudness. *Proc Natl Acad Sci U S A, 54*(2), 407-411.

Stevens, S.S. (1957). On the psychophysical law. *Psychol Rev, 64*(3), 153-181.

Stevenson, R.A., & James, T.W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. Neuroimage, 44(3), 1210-1223.

Stiles, S.W. (1959). Induction of ether anesthesia in cats. *J Am Vet Med Assoc, 134*(6), 275.

Stilla, R., & Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Hum Brain Mapp, 29*(10), 1123-1138.

Stockman, A., Sharpe, L.T., Zrenner, E., & Nordby, K. (1991). Slow and fast pathways in the human rod visual system: electrophysiology and psychophysics. *J Opt Soc Am A, 8*(10), 1657-1665.

Stoffels, E.J., & van der Molen, M.W. (1988). Effects of visual and auditory noise on visual choice reaction time in a continuous-flow paradigm. *Percept Psychophys, 44(1),* 7-14.

Stoffels, E.J., van der Molen, M.W., & Keuss, P.J. (1989). An additive factors analysis of the effect(s) of location cues associated with auditory stimuli on stages of information processing. *Acta Psychol (Amst), 70(2)*, 161-197.

Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J Neurosci, 26(43), 11138-11147.

Sumner, P. (2006). Inhibition versus attentional momentum in cortical and collicular mechanisms of IOR. *Cogn Neuropsychol, 23*(7), 1035-1048.

Sumner, P., Adamjee, T., & Mollon, J.D. (2002). Signals invisible to the collicular and magnocellular pathways can capture visual attention. *Curr Biol, 12*(15), 1312-1316.

Sumner, P., Nachev, P., Vora, N., Husain, M., & Kennard, C. (2004). Distinct cortical and collicular mechanisms of inhibition of return revealed with S cone stimuli. *Curr Biol, 14*(24), 2259-2263.

Tal, N., & Amedi, A. (2009). Multisensory visual-tactile object related network in humans: insights gained using a novel crossmodal adaptation approach. *Exp Brain Res, 198*(2-3), 165-182.

Talsma, D., Doty, T.J., & Woldorff, M.G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex, 17(3),* 679-690.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M.G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci, 14*(9), 400-410.

Tardif, E., Delacuisine, B., Probst, A., Clarke, S. (2005) Intrinsic connectivity of human superior colliculus. *Exp Brain Res 166*:316-324.

Taylor, K.I., Moss, H.E., Stamatakis, E.A., & Tyler, L.K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc Natl Acad Sci U S A, 103*(21), 8239-8244.

Todd, J. W. (1912). *Reaction to Multiple Stimuli.* New York: The Science Press.

Trappenberg, T.P., Dorris, M.C., Munoz, D.P., & Klein, R.M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. Journal of Cognitive Neuroscience, 13(2), 256–271.

Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn Psychol, 12(1),* 97-136.

Ulrich, R., Miller, J., & Schroter, H. (2007). Testing the race model inequality: an algorithm and computer programs. *Behav Res Methods, 39*(2), 291-302.

Ulrich, R., Rinkenauer, G., & Miller, J. (1998). Effects of stimulus duration and intensity on simple reaction time and response force. *J Exp Psychol Hum Percept Perform, 24*, 915–928.

Van der Burg, E., Olivers, C.N., Bronkhorst, A.W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform, 34*(5), 1053-1065.

van der Smagt, M.J., van Engeland, H., & Kemner, C. (2007). Brief report: can you see what is not there? low-level auditory-visual integration in autism spectrum disorder. *J Autism Dev Disord, 37*(10), 2014-2019.

van Ee, R., van Boxtel, J.J., Parker, A.L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *J Neurosci, 29*(37), 11641-11649.

Vergara, M., Mondragon, S., Sancho-Bru, J.L., Company, P., & Agost, M.J. (2011). Perception of products by progressive multisensory integration. A study on hammers. *Appl Ergon, 42*(5), 652-664.

Vitu, F., Lancelin, D., Jean, A., & Farioli, F. (2006). Influence of foveal distractors on saccadic eye movements: a dead zone for the global effect. *Vision Res, 46*(28), 4684-4708.

von Stein, A., Rappelsberger, P., Sarnthein, J., & Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cereb Cortex, 9*(2), 137-150.

Vroomen, J. (1999). Ventriloquism and the nature of the unity assumption. In G. Aschersleben, T. Bachmann & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 388-394). North-Holland: Elsevier.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychol (Amst), 108*(1), 21-33.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept Psychophys, 63(4),* 651-659.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). Directing spatial attention towards the illusory location of a ventriloquized sound. Acta Psychol (Amst), 108(1), 21-33.

Wald, G. (1966). Defective color vision and its inheritance. *Proc Natl Acad Sci U S A, 55*(6), 1347-1363.

Walker, P., & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception, 14*(6), 729-736.

Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S.P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychol Sci, 21*(1), 21-25.

Walker, R., Deubel, H., Schneider, W.X., & Findlay, J.M. (1997). Effect of remote distractors on saccade programming: evidence for an extended fixation zone. J Neurophysiol, *78*:1108–1119

Walker, R., Kentridge, R.W., & Findlay, J.M. (1995). Independent contributions of the orienting of attention, fxation onset and bilateral stimulation on human saccadic latency. *Exp Brain Res, 103*:294– 310.

Walker, R., Mannan, S., Maurer, D., Pambakian, A.L., & Kennard, C. (2000). The oculomotor distractor effect in normal & hemianopic vision. Proc Biol Sci, *7*;267(1442):431-8

Wallace, M.T., & Stein, B.E. (1994). Cross-modal synthesis in the midbrain depends on input from cortex. *J Neurophysiol, 71*(1), 429-432.

Wallace, M.T., & Stein, B.E. (1996). Sensory organization of the superior colliculus in cat and monkey. *Prog Brain Res, 112*, 301-311.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol, 69*(6), 1797-1809.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1998). Multisensory integration in the superior colliculus of the alert cat. *J Neurophysiol, 80*(2), 1006-1010.

Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol, 76*(2), 1246-1266.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *TRENDS in Cognitive Sciences, 7*, 483-488.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-color synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex, 42*, 264–280.

Welch, R.B. (1999a). Meaning, attention, and the ‘‘unity assumption” in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T.Bachmann & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events.* Amsterdam: Elsevier.

Welch, R.B. (1999b). Meaning, sttention and the "unity assumption" in the intersensory bia of spatial and temporal perceptions. In G.Aschersleben, T.Bachmann & J.Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (Vol. 129, pp. 371-387). Amsterdam: Elsevie.

Welch, R.B., & Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol Bull, 88*, 638-667.

White, B.J., Kerzel, D., & Gegenfurtner, K. R. (2006). Visually guided movements to color targets. *Exp Brain Res, 175*(1), 110-126.

Wilkinson, L.K., Meredith, M.A., & Stein, B.E. (1996). The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp Brain Res, 112*(1), 1-10.

Wurtz, R.H., & Goldberg, M.E (Eds.) (1989). The Neurobiologv of Saccadic Eve Movements. Amsterdam: Elsevier

Wurtz, R.H., Sommer, M.A., Paré, M., & Ferraina, S. (2001). Signal transformations from cerebral cortex to superior colliculus for the generation of saccades. *Vis. Research, 41*: 3399-3412.

Wyszecki, G. (1967). Correlate for lightness in terms of CIE chromaticity coordinates and luminous reflectance. *J Opt Soc Am., 57*:254–257.

Yao, L. & Peck, C.K. (1997) Saccadic eye movements to visual and auditory targets. *Exp. Brain Res., 115*: 25-34.

Zahn, J.R., Abel, L.A., & Dell'Oso, L.F. (1978). The audio-ocular response characteristics. *Sensory Processes, 2*, 32–37.

Zambarbieri, D., Andrei, I., Beltrami, G., Fontana, L. & Versino, M. (1995 a). Saccade latency towards auditory targets. In J.M. Findlay, R. Walker, and R.W. Kentridge (eds.) Eye Movement Research: Mechanisms, Processes and Applications (pp. 157 – 164). Amsterdam: Elsevier Science. 27-33.

Zambarbieri, D., Schmid, R., Magenes, G., & Prablanc, C. (1982). Saccadic responses evoked by presentation of visual and auditory targets. *Experimental Brain Research*, *47*, 417–427.

Zambarbieri, D., Schmid, R., Prablanc, C., & Magenes, G. (1981). Characteristics of eye movements evoked by the presentation of acoustic targets. In A.F. Fuchs and W. Becker (eds.). Progress in oculomotor research (pp.559 – 566). Amsterdam: Elsevier/North Holland.

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature, 335*(6188), 311-317.

APPENDIX

This appendix provides a specification of the steps I performed for testing the violations of the race model (RM: Miller, 1982; Ulrich, Miller & Schröter, 2007) and other aspects of data analysis adopted in my thesis.

Using data from one representative participant (Table 1) the analysis is broken up in six steps.

Table 1: Illustrative sets of RTs (in ms) for each experimental condition for a single representative participant in Experiment 6.

Condition	*n*	*Data set of RTs*											
A	***32***	317	326	331	335	346	347	.....	513	518	525	532	532
V	***38***	327	334	348	352	352	359	.....	459	465	472	496	522
MSC	***39***	260	276	278	284	292	293	.....	354	371	379	382	384

Step 1: Empirical cumulative probability distribution for every participant in each experimental condition.

For each condition (A, V and MSC) the RT set has been ordered from the smallest to the largest value, and cumulative probability distributions (CPD) have been calculated with a 1-ms bin resolution (Figure 1A).

Step 2: Computation of P (RM) = P (A) + P (V)

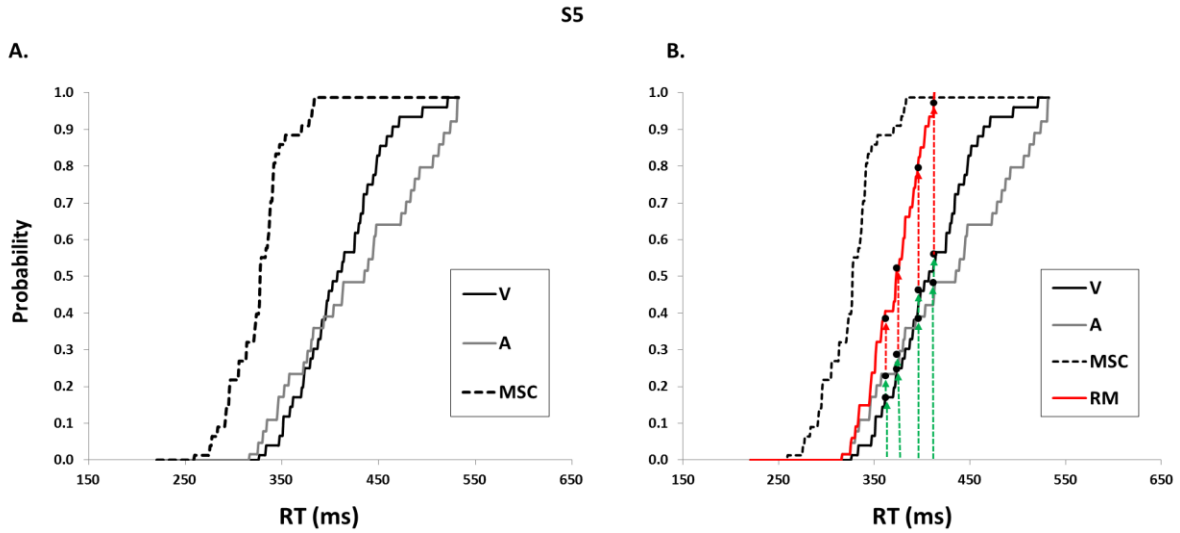


Figure 1. A) Empirical CPD for participant 5 in the three experimental conditions: visual target only (curvy, V); auditory (soft "wow") target only (A), multisensory congruent condition for the curvy + wow combination (MSC). B) for every RT, the corresponding values of P(A) and P(V) are summed in order to obtain the probability corresponding to the upper limit of the Race model. The resulting CPD of the RM is shown in red. In this case, for each RT the MSC value is higher than the maximum value predicted by the race model, indicating a clear violation of additivity.

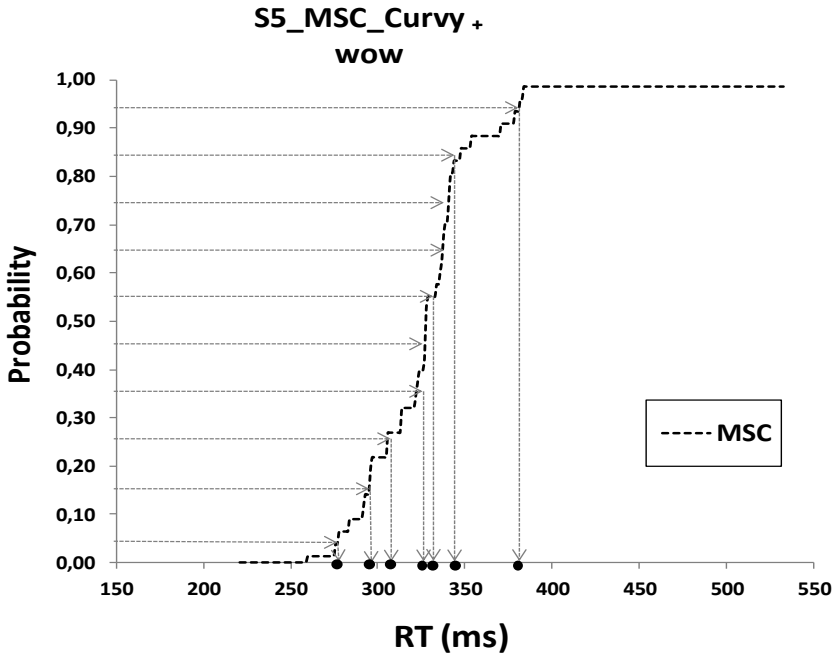
The probabilities for conditions A and V were added in order to obtain the upper limit of the distribution predicted by the RM. The comparison of CPDs for the MSC condition and the upper limit of the RM defines the violation of additivity for a given participant.

Step 3: Statistical analysis

In order to test for the statistical significance of violations in a group of participants, different approaches are adopted in the literature. The first procedure, described in detail by Ulrich et al. (2007), consists in a series of t tests, one for each predefined percentile (see also Bucur, Allen, Sanders, Ruthruff, & Murphy, 2005; Krummenacher, Müller, & Heller, 2001). The second consists in a geometrical measure of the strength of the violation (Miller, 1986; Colonius & Diederich, 2006). Both approaches have been adopted in my thesis.

Table 2. Illustrative sets of empirical CPDs for A and V conditions (1-ms resolution) and the CPD representing the upper limit of the RM, obtained from the sum of P(A) and P(V).

	Empirical CPD				Predicted CPD
ms	P (A)		P (V)		P (RM)
221	0,000	+	0,000	→	0,000
222	0,000	+	0,000	→	0,000
223	0,000	+	0,000	→	0,000
224	0,000	+	0,000	→	0,000
225	0,000	+	0,000	→	0,000
.....	.....		.....		.....
359	0,145	+	0,234	→	0,379
360	0,145	+	0,234	→	0,379
361	0,158	+	0,234	→	0,392
362	0,171	+	0,234	→	0,405
363	0,171	+	0,234	→	0,405
364	0,171	+	0,234	→	0,405
.....	.....		.....		.....
400	0,461	+	0,391	→	0,851
401	0,461	+	0,391	→	0,851
402	0,474	+	0,391	→	0,864
403	0,487	+	0,406	→	0,893
404	0,487	+	0,422	→	0,909
405	0,487	+	0,422	→	0,909
.....	.....		.....		.....
411	0,513	+	0,422	→	0,935
412	0,526	+	0,438	→	0,935
413	0,539	+	0,469	→	0,993
414	0,553	+	0,484	→	1,024
415	0,566	+	0,484	→	1,050
416	0,566	+	0,484	→	1,050



p	MSC	RM_MSC
0,05	237	326
0,15	282	345
0,25	308	351
0,35	327	358
0,45	333	372
0,55	374	379
0,65	392	382
0,75	409	393
0,85	419	399
0,95	496	412

Figure 2. For each predefined percentile (0.05; 0.15; 0.25; 0.35; 0.45; 0.55; 0.65; 0.75; 0.85; 0.95) a RT *value* was derived (see black dots, upper graph) for each empirical cumulative distribution (A,V and MSC) and the predicted cumulative distribution of the RM (see lower table for numbers relative to one subject). This procedure, here represented for one single subject, was performed for each subject separately.

Figure 2 illustrate the procedure followed for the implementation of the first method of analysis of the Race model. For each predefined percentile the corresponding *RT value* was obtained, for the percentiles 0.05; 0.15; 0.25; 0.35; 0.45; 0.55; 0.65; 0.75; 0.85; 0.95. This procedure was followed for each MS condition (congruent, incongruent, spiky or curvy) and the corresponding Race model predictions (see Figure 3, left table) and for each subject separately. In this way, I obtained a series of *RT* values (one for each subject) for each percentile of the distribution MS and RM distributions, for each subject separately. The difference in *RT* values were calculated between MS and RM at each percentile for each subject, and such a series of *RT differences* of every subject were tested against 0 for each percentile separately (see Table 3).

Table 3. Illustrative set of differences between RT values (in ms; corresponding to predefined percentiles) for the RM_MSC –MSC conditions. Each series of RT differences were tested against 0 for every percentile separately.

	0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
S1	28	36	30	32	22	14	8	-1	-10	-72
S2	18	18	21	16	11	12	17	16	19	20
S3	48	35	38	38	29	21	14	12	-4	-19
S4	27	6	12	13	13	8	-2	1	-16	-23
S5	49	51	45	36	44	50	45	52	52	32
S6	21	15	10	11	9	0	-13	-16	-25	-54
S7	55	49	40	49	43	36	37	25	22	-2
S8	68	54	47	40	38	35	26	18	13	12
Mean	**39**	**33**	**30**	**29**	**26**	**22**	**17**	**13**	**6**	**-14**
s.e.m.	**6.9**	**6.9**	**5.5**	**5.3**	**5.5**	**6.4**	**7.3**	**7.7**	**9.5**	**13.6**

The mean RT differences for each percentile were plotted relative to the percentiles of the distribution (Figure 3).

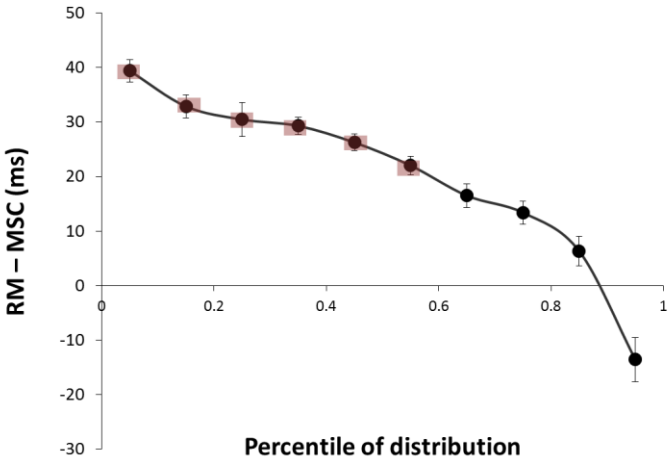


Figure 3. The mean difference of RT values (and s.e.m) between RT_MSC and RT_RM at each percentile are plotted in relation to the percentiles of the distribution. A significant difference indicates a violation of the race inequality test for the MSC condition. The red rectangles mark the areas in which the violation is significantly different from zero, as assessed by a series of one-tailed t-tests (corrected fo multiple comparisons).

The alternative procedure is described in the following section.

Area of violation

If many different experimental conditions are to be compared with respect to their degree of RM violation (for example, the violation obtained by MSC vs. MSI conditions), it is desirable to reduce the violation to a single index (Colonius & Diederich, 2006; Hughes et al., 1998), typically the area within the contours defined by the multisensory condition and the RM.

This geometrical measure was introduced by Miller (1986) in order to assess the degree of violation of the inequality, and was later adopted also by other authors (e.g. Colonius & Diederich, 2006). The area of violation is typically interpreted as the strength of neural summation or co-activation.

However, when distribution properties have to be compared across different participants it is opportune to normalize each distribution. In order to achieve this, each

participant's response times were divided into 21 quantiles; i.e., the range between 0 and 100% was quantised with a resolution of 5% (Figure 4a).

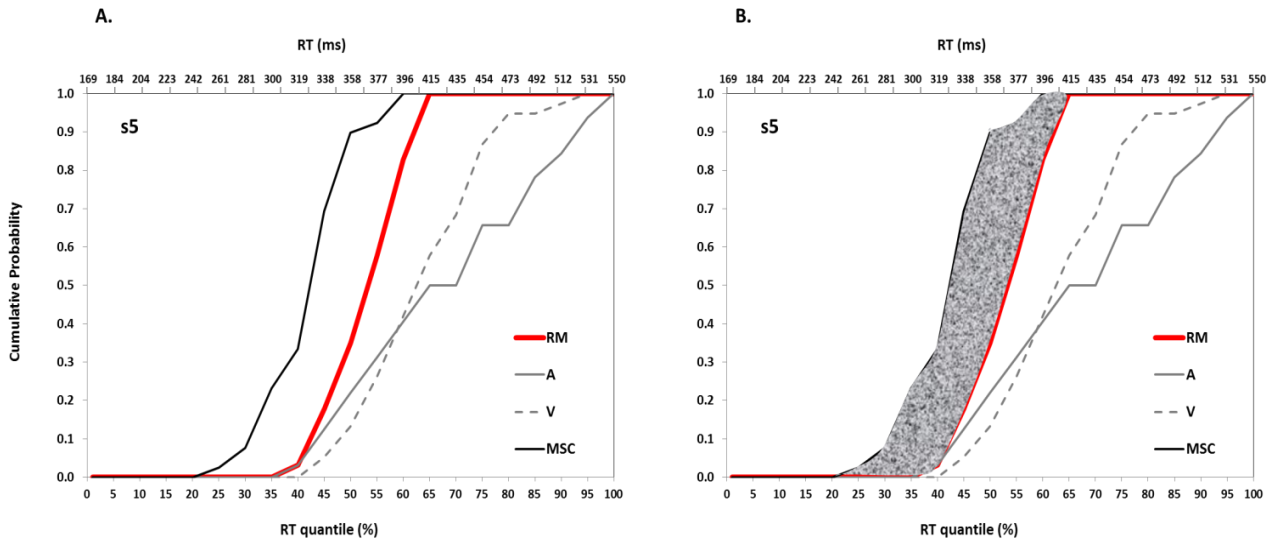


Figure 4. A) Empirical cumulative probabilities for A, V and MSC conditions and cumulative probabilities predicted by the RM (A+V), as a function of raw RTs (upper abscissa) and 5% quantiles (lower abscissa). **B)** Illustrative representation of the "area of violation" calculated for participant 5.

The positive area between the CDP of the multisensory MSC condition and the upper boundary of the RM was calculated for each single participant (as shown in Figure 4 B for the normalized distributions of participant 5), and entered into a Shape x Congruency ANOVA (see main text for the analysis).

Race model of the inhibition

Since the effect studied in Chapter 5 concerns the strength of inhibition rather than enhancement, the predicted boundary for the Race model were obtained for an inhibiting effect of AV distractors. In order to achieve this, the following reasoning was adopted:

a) The CPDs were derived for each experimental condition: target only (T), visual distractor (T+VD), auditory distractor (T+AD) and audio-visual distractor (T+AVD) (Figure 5a, Table 3a)

b) The difference profile relative to the target only baseline was derived for each unisensory distractor (either A or V; Table 4a), by subtracting at every quantile the CPD of the target condition from each CPD of distractor condition (p_T+VD; p_T+AD) (Figure 5b, Table 4b)

$$A/V_{\text{inhibition}} = P_{\text{distractor (A,V)}} - P_{\text{target}}$$

c) The maximum amount of inhibition produced by AVD according to the additive combination of A and V inhibitions was calculated ($A_{\text{inhibition}} + V_{\text{inhibition}}$, see Figure 5b, Table 4c)

$$\text{Strength of inhibition\_AV} = A_{\text{inhibition}} + V_{\text{inhibition}}$$

d) the predicted value of the inhibition according to the RM was derived adding the strength of AV inhibition to the CDP of the target condition (Figure 5c, Table 4d)

$$RM\__{inhibition} = P_{\text{target}} + \text{Strength of inhibition\_AV}$$

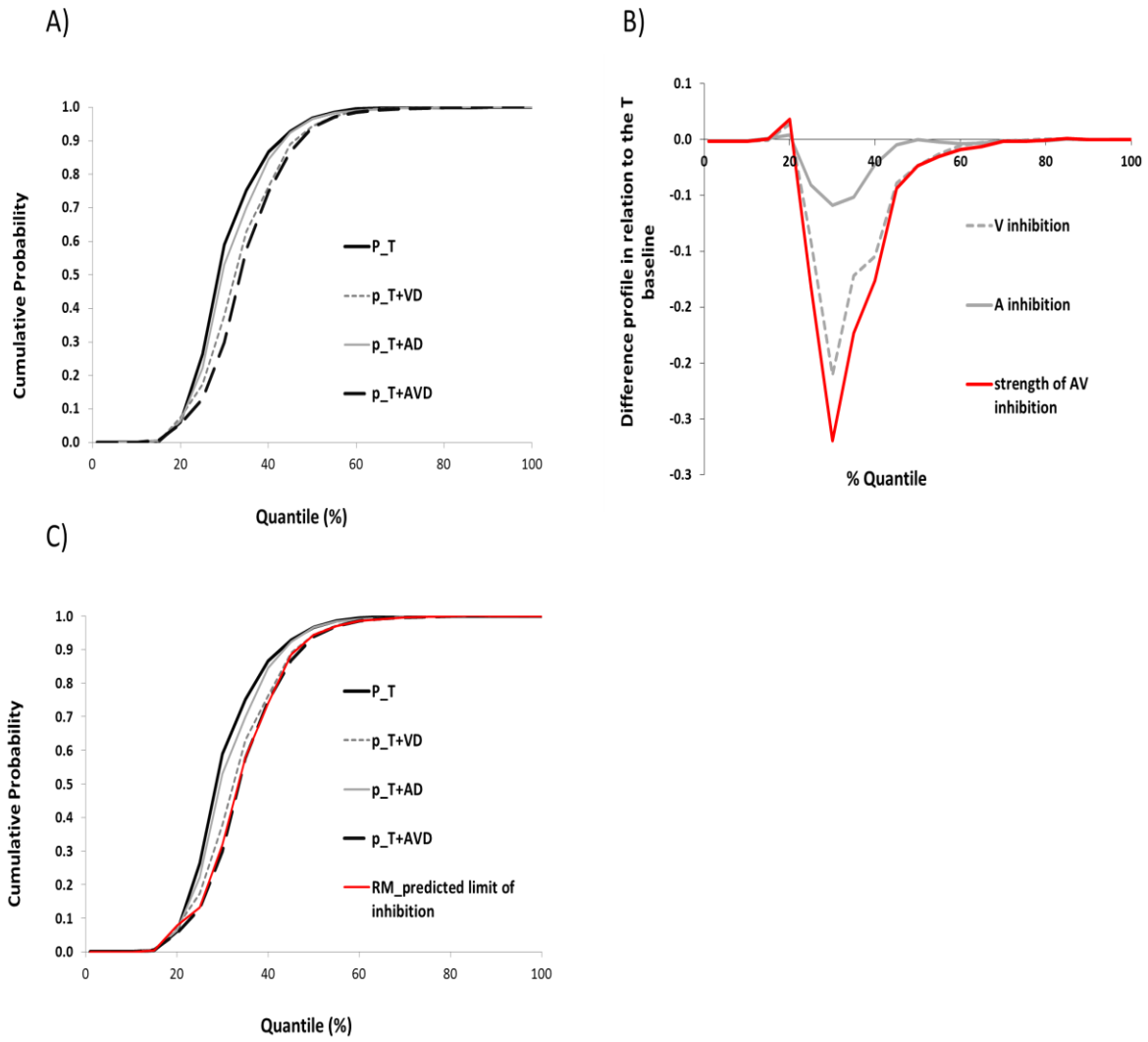


Figure 5. A) Illustrative representation of empirical CPDs for the target only (T) baseline, target + visual distractor (T+VD), target + auditory distractor (T+AD) and target + audio-visual distractor (T+AVD) of one representative participant in Experiment 8. **B)** The profile of the % of change in the two distractor condition (A or V; see also Table 4b) relative to the T baseline. The red line represents the derived strength of the AV inhibition, obtained adding the A_inhibition to V_inhibition profile (Table 4c). **D)** Predicted CPD of the RM of the inhibition (red line) where obtained adding the CPD of the target to the predicted strength of inhibition profile for AV (Table 4d). The graph shows no evidence of the violation of the RM (the AVD distribution and the RM_inhibition distributions shows perfect overlapping).

Table 4. **Illustrative sets of the empirical CPDs obtained for T, AD and VD conditions and of the predicted RM of the inhibition. The data are represented following the steps described in the main text and in Figure 5.**

a)				b)		c)		d)
p T	*p*_T+VD	*p*_T+AD		V *inhib* (p_T+VD - p_T)	A *inhib* (p_T+AD -p_T)	*Strength of inhibition* AV (V*inhib* + A*inhib*)		RM_inhibition (p_T + strength of inhibition AV)
0.00	0.00	0.00	→	-0.001	-0.001	-0.002	→	0.00
0.00	0.00	0.00	→	-0.001	-0.001	-0.002	→	0.00
0.00	0.00	0.00	→	-0.001	-0.001	-0.002	→	0.00
0.00	0.00	0.01	→	0.000	0.001	0.001	→	0.01
0.06	0.08	0.07	→	0.014	0.004	0.018	→	0.08
0.26	0.17	0.22	→	-0.090	-0.041	-0.131	→	0.13
0.59	0.38	0.53	→	-0.211	-0.059	-0.269	→	0.32
0.75	0.63	0.70	→	-0.122	-0.051	-0.173	→	0.58
0.87	0.76	0.84	→	-0.104	-0.022	-0.127	→	0.74
0.93	0.89	0.92	→	-0.039	-0.005	-0.044	→	0.88
0.97	0.94	0.97	→	-0.024	0.000	-0.024	→	0.94
0.99	0.97	0.98	→	-0.013	-0.002	-0.015	→	0.97
0.99	0.99	0.99	→	-0.005	-0.004	-0.009	→	0.99
1.00	0.99	0.99	→	-0.003	-0.003	-0.006	→	0.99
1.00	1.00	1.00	→	-0.001	-0.001	-0.002	→	1.00
1.00	1.00	1.00	→	0.000	-0.001	-0.001	→	1.00
1.00	1.00	1.00	→	0.000	-0.001	0.000	→	1.00
1.00	1.00	1.00	→	0.000	0.000	0.001	→	1.00
1.00	1.00	1.00	→	0.000	0.000	0.000	→	1.00
1.00	1.00	1.00	→	0.000	0.000	0.000	→	1.00
1.00	1.00	1.00	→	0.000	0.000	0.000	→	1.00